ARO 26394.2-MA

2

ASYMPTOTIC ROBUSTNESS IN REGRESSION AND AUTOREGRESSION BASED ON LINDEBERG CONDITIONS

T. W. Anderson and Naoto Kunitomo

Stanford University

TECHNICAL REPORT NO. 23

June 1989

U. S. Army Research Office

Contract DAAL03-89-K-0033

Theodore W. Anderson, Project Director

Department of Statistics

Stanford University

Stanford, California

ASYMPTOTIC ROBUSTNESS IN REGRESSION AND AUTOREGRESSION BASED ON LINDEBERG CONDITIONS

T. W. Anderson and Naoto Kunitomo
Stanford University

TECHNICAL REPORT NO. 23
June 1989

U. S. Army Research Office
Contract DAAL03-89-K-0033
Theodore W. Anderson, Project Director

Department of Statistics
Stanford University
Stanford, California

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO 26394.2-MA	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) ASYMPTOTIC ROBUSTNESS IN REGRESSION AND AUTOREGRESSION BASED ON LINDEBERG CONDITIONS		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) T. W. Anderson and Naoto Kunitomo		8. CONTRACT OR GRANT NUMBER(s) DAAL03-89-K-0033
9. PERFORMING ORGANIZATION NAME AND ADDRESS Stanford University Department of Statistics - Sequoia Hall Stanford, California 94305-4065		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE June 1989
		13. NUMBER OF PAGES 31pp.
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Asymptotic robust, Lindeberg condition, central limit theorem, regression coefficients, autoregression coefficients

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See reverse side for abstract.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract.

A statistical procedure is asymptotically robust if its large-sample properties hold under conditions more general than the conditions under which the procedure is derived. The justification of such properties is often based directly or indirectly on a central limit theorem. In this paper a form of the Lindeberg condition appropriate for martingale differences is used to obtain consistency and asymptotic normality of statistics for regression and autoregression. The regression model is $\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{z}_t + \boldsymbol{v}_t$. The unobserved error sequence $\{\boldsymbol{v}_t\}$ is a sequence of martingale differences with conditional covariance matrices $\{\boldsymbol{\Sigma}_t\}$ and satisfying

$$\frac{1}{n} \sup_{t=1,\dots,n} \mathcal{E}\left\{ v_t' v_t I(v_t' v_t > a) \middle| \boldsymbol{z}_t, \boldsymbol{v}_{t-1}, \boldsymbol{z}_{t-1}, \dots \right\} \xrightarrow{p} 0$$

as $a \to \infty$. The sample covariance of the independent variables, $\boldsymbol{z}_1, \dots, \boldsymbol{z}_n$, is assumed to have a probability limit $\boldsymbol{M}$, constant and nonsingular; $\max_{t=1,\dots,n} \boldsymbol{z}_t' \boldsymbol{z}_t / n \xrightarrow{p} 0$. If $(1/n)\sum_{t=1}^{n} \boldsymbol{\Sigma}_t \xrightarrow{p} \boldsymbol{\Sigma}$, constant, then $\sqrt{n}\,\mathrm{vec}(\widehat{\boldsymbol{B}}_n - \boldsymbol{B}) \xrightarrow{\mathcal{L}} N(0, \boldsymbol{M}^{-1} \otimes \boldsymbol{\Sigma})$.

The autoregression model is $\boldsymbol{x}_t = \boldsymbol{B}\boldsymbol{x}_{t-1} + \boldsymbol{v}_t$ with the above conditions on $\{\boldsymbol{v}_t\}$ and

$$\frac{1}{n} \sum_{t=\max(r,s)+1}^{n} (\boldsymbol{\Sigma}_t \otimes \boldsymbol{v}_{t-1-r} \boldsymbol{v}_{t-1-s}') \xrightarrow{p} \delta_{rs} (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}),$$

where δ_{rs} is the Kronecker delta. Then $\sqrt{n}\,\mathrm{vec}(\widehat{\boldsymbol{B}}_n - \boldsymbol{B}) \xrightarrow{\mathcal{L}} N(\boldsymbol{0}, \boldsymbol{\Gamma}^{-1} \otimes \boldsymbol{\Sigma})$, where $\boldsymbol{\Gamma} = \sum_{s=0}^{\infty} \boldsymbol{B}^s \boldsymbol{\Sigma} (\boldsymbol{B}')^s$.

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

1. Introduction.

A statistical procedure is asymptotically robust if its large-sample properties hold under conditions more general than the conditions under which the procedure is derived. The justification of such procedures is often based directly or indirectly on a central limit theorem. In this paper Lindeberg-type conditions are utilized to establish asymptotic normality of sample regression and autoregression coefficients.

The classic central limit theorem for independent identically distributed scalar random variables $x_1, x_2, \ldots$ states that $\sqrt{n}\ \bar{x}_n \xrightarrow{\mathcal{L}} N(0, \sigma^2)$ as $n \to \infty$ if $\mathcal{E}x_i = 0$ and $\mathcal{E}x_i^2 = \sigma^2$; here $\bar{x}_n = \sum_{i=1}^n x_i/n$ is the mean of the first n observations. The requirement that the variables be identically distributed can be dropped. For $\mathcal{E}x_i = 0$ and $\mathcal{E}x_i^2 = \sigma_i^2$,

(1.1) $$\frac{1}{\tau_n}\sum_{i=1}^n x_i \xrightarrow{\mathcal{L}} N(0,1),$$

where

(1.2) $$\tau_n^2 = \sum_{i=1}^n \sigma_i^2,$$

if for any given $\varepsilon > 0$

(1.3) $$\frac{1}{\tau_n^2}\sum_{i=1}^n \mathcal{E}x_i^2 I(x_i^2 > \varepsilon\tau_n^2) \longrightarrow 0$$

as $n \to \infty$. Here $I(\cdot)$ is the indicator function. If $\sigma_n^2/\tau_n^2 \to 0$ as $n \to \infty$, then (1.1) implies (1.3); in this sense the Lindeberg (1922) condition (1.3) is minimal.

The condition of independence can be weakened to a condition of martingale differences. A very general theorem, which we shall use, has been given by Dvoretzky (1972). For justification of later theorems we state this result in terms of a triangular array of random variables (and include a normalization in the definition of the random variables).

Theorem (Dvoretzky). Let $x_{n1}, \ldots, x_{nn}$ be a set of random variables and $\mathcal{F}_{n0} \subset \mathcal{F}_{n1} \subset \cdots \subset \mathcal{F}_{nn}$ be a set of σ-fields, $n = 1, 2, \ldots$, such that x_{nj} is $\mathcal{F}_{nj}$-measurable.

(1.4) $$\mathcal{E}(x_{nj}|\mathcal{F}_{n,j-1}) = 0 \quad \text{a.s.},$$

(1.5) $$\mathcal{E}(x_{nj}^2|\mathcal{F}_{n,j-1}) = \sigma_{nj}^2 \quad \text{a.s.},$$

$$\sum_{i=1}^{n} \sigma_{ni}^2 \xrightarrow{\mathrm{P}} \sigma^2 \tag{1.6}$$

as $n \to 0$, where σ^2 is constant, and for any given $\varepsilon > 0$

$$\sum_{t=1}^{n} \mathcal{E}\left[x_{nj}^2 I(x_{nj}^2 > \varepsilon) \middle| \mathcal{F}_{n,j-1}\right] \xrightarrow{\mathrm{P}} 0. \tag{1.7}$$

Then

$$\sum_{j=1}^{n} x_{nj} \xrightarrow{\mathcal{L}} N(0, \sigma^2). \tag{1.8}$$

Dvoretzky actually showed that this result holds if $\mathcal{F}_{n,j-1}$ is replaced by $\mathcal{B}_{n,j-1}$, the σ-field generated by $\sum_{i=1}^{j-1} x_{ni}$. Generalizations have been given in Section 3.2 of Hall and Heyde (1980) and Section 9.5 of Chow and Teicher (1988). Further references can be found in these books.

In this paper we consider the estimation of the matrix of regression coefficients $\boldsymbol{B}$ in the model

$$\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{z}_t + \boldsymbol{v}_t, \quad t = 1, 2, \ldots\ , \tag{1.9}$$

where the unobservable vector disturbances $\boldsymbol{v}_t$ are martingale differences; that is, the conditional expected value of $\boldsymbol{v}_t$ given earlier observed $\boldsymbol{y}_t$'s and $\boldsymbol{z}_t$'s is $\mathbf{0}$. The conditional second-order moments of the $\boldsymbol{v}_t$'s are finite, but not necessarily the same for all t. However, the $\boldsymbol{v}_t$'s satisfy a kind of Lindeberg condition. The "independent" variables $\boldsymbol{z}_t$ are assumed to have a sample covariance matrix that converges to a limit in probability, and the $\boldsymbol{z}_t$'s satisfy a kind of asymptotic negligibility condition. It is shown that the least squares estimator of $\boldsymbol{B}$ has an asymptotic distribution that is the same as in the case that the $\boldsymbol{v}_t$'s are independent and normal with mean $\mathbf{0}$ and constant covariance matrix. Thus the disturbances do not need to be homoscedastic nor do they need to be independent. The relaxed conditions are particularly important when the observed $\boldsymbol{z}_t$'s and $\boldsymbol{y}_t$'s constitute a time series.

In the autoregressive model, which is extensively used in time series analysis,

$$\boldsymbol{x}_t = \boldsymbol{B}\boldsymbol{x}_{t-1} + \boldsymbol{v}_t, \quad t = 1, 2, \ldots\ , \tag{1.10}$$

the vector z_t is replaced by x_{t-1}. The conditions on the v_t's imply the desired conditions on the x_{t-1}'s.

In Section 4 the mixed model is considered; the right-hand side may contain both lagged "dependent" variables and independent variables.

If the disturbances in the regression model are normal, independent, and homoscedastic, and the independent variables are nonstochastic, the estimator of B has a normal distribution with expected value B and covariances determined by the common covariance matrix of the disturbances; it follows that the asymptotic distribution is normal. The restriction of homoscedasticity was relaxed by Anderson (1971) in Theorems 2.6.1 and 2.6.2 under a Lindeberg-type condition on the disturbances and the condition that the sample covariance matrix of the independent variables have a nonsingular limit.

In the autoregression model the least squares estimator of B is nonlinear in the disturbances. Mann and Wald (1943) showed that the asymptotic distribution of the estimator of B is normal under the condition that the disturbances are independently identically distributed and possess moments of all orders. Anderson (1959) showed that in this case *only the second-order moments need to be finite.*

There are many recent results in this area. Lai and Robbins (1981) proved a theorem for a scalar dependent variable with independent identically distributed disturbances. Lai and Wei (1982) proved a similar theorem under the conditions that the moments of the disturbances of some order greater than 2 are bounded and that the variances of the disturbances converge to a constant a.s. Our approach follows these papers, but the conditions have been relaxed. Chan and Wei (1987) have used a Lindeberg condition for a special case of the autoregressive process; see also Lai and Siegmund (1983).

2. Robustness in Regression.

We consider the regression model in which the observed vector-valued dependent variable y_t is generated by

$$y_t = Bz_t + v_t, \quad t = 1, 2, \dots , \tag{2.1}$$

where z_t is an observed vector-valued independent variable and $\{v_t\}$ is a sequence of (unobservable) martingale differences satisfying a Lindeberg-type condition.

Theorem 1. Let $\{\boldsymbol{z}_t, \boldsymbol{v}_t\}$, $t = 1, 2, \ldots$, be a sequence of pairs of random vectors, and let $\{\mathcal{F}_t\}$ be an increasing sequence of σ–fields such that $\boldsymbol{z}_t$ is $\mathcal{F}_{t-1}$-measurable and $\boldsymbol{v}_t$ is $\mathcal{F}_t$–measurable. Let the matrix $\boldsymbol{D}_n$ be $\mathcal{F}_0$–measurable such that

$$\boldsymbol{D}_n^{-1} \sum_{t=1}^{n} \boldsymbol{z}_t \boldsymbol{z}_t' (\boldsymbol{D}_n')^{-1} \xrightarrow{\mathrm{p}} \boldsymbol{C}, \tag{2.2}$$

a constant matrix, as $n \to \infty$, and

$$\max_{t=1,\ldots,n} \boldsymbol{z}_t' (\boldsymbol{D}_n \boldsymbol{D}_n')^{-1} \boldsymbol{z}_t \xrightarrow{\mathrm{p}} 0. \tag{2.3}$$

Suppose further that $\mathcal{E}(\boldsymbol{v}_t | \mathcal{F}_{t-1}) = \boldsymbol{0}$ a.s., $\mathcal{E}(\boldsymbol{v}_t \boldsymbol{v}_t' | \mathcal{F}_{t-1}) = \boldsymbol{\Sigma}_t$ a.s.,

$$\sum_{t=1}^{n} \left[\boldsymbol{\Sigma}_t \otimes \boldsymbol{D}_n^{-1} \boldsymbol{z}_t \boldsymbol{z}_t' (\boldsymbol{D}_n')^{-1}\right] \xrightarrow{\mathrm{p}} \boldsymbol{\Sigma} \otimes \boldsymbol{C}, \tag{2.4}$$

where $\boldsymbol{\Sigma}$ is a constant positive semidefinite matrix, and

$$\sup_{t=1,2,\ldots} \mathcal{E}\left[\boldsymbol{v}_t' \boldsymbol{v}_t I(\boldsymbol{v}_t' \boldsymbol{v}_t > a) | \mathcal{F}_{t-1}\right] \xrightarrow{\mathrm{p}} 0 \tag{2.5}$$

as $a \to \infty$. Then

$$\operatorname{vec}\left(\boldsymbol{D}_n^{-1} \sum_{t=1}^{n} \boldsymbol{z}_t \boldsymbol{v}_t'\right) \xrightarrow{\mathcal{L}} N(\boldsymbol{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{C}). \tag{2.6}$$

Proof. The conclusion holds if

$$\begin{aligned} \operatorname{tr} \boldsymbol{D}_n^{-1} \sum_{t=1}^{n} \boldsymbol{z}_t \boldsymbol{v}_t' \boldsymbol{B} &= \sum_{t=1}^{n} \boldsymbol{v}_t' \boldsymbol{B} \boldsymbol{D}_n^{-1} \boldsymbol{z}_t \\ &\xrightarrow{\mathcal{L}} N(\boldsymbol{0}, \operatorname{tr} \boldsymbol{\Sigma B C B}') \end{aligned} \tag{2.7}$$

for every $\boldsymbol{B}$. Let $\boldsymbol{u}_{nt} = \boldsymbol{B D}_n^{-1} \boldsymbol{z}_t$, $t = 1, \ldots, n$. Then

$$\sum_{t=1}^{n} \boldsymbol{u}_{nt} \boldsymbol{u}_{nt}' \xrightarrow{\mathrm{p}} \boldsymbol{BCB}' = \boldsymbol{D}, \tag{2.8}$$

say. We want to show that

$$\sum_{t=1}^{n} \boldsymbol{u}_{nt}' \boldsymbol{v}_t \xrightarrow{\mathcal{L}} N(0, \operatorname{tr} \boldsymbol{\Sigma D}). \tag{2.9}$$

Condition (2.3) implies

$$\max_{t=1,\ldots,n} \boldsymbol{u}'_{nt}\boldsymbol{u}_{nt} \xrightarrow{\text{p}} 0. \tag{2.10}$$

Let

$$\boldsymbol{w}_{nt} = \boldsymbol{u}_{nt} I(\|\boldsymbol{u}_{nt}\| \leq 1), \quad t = 1, \ldots, n, \ n = 1, 2, \ldots. \tag{2.11}$$

Then $\|\boldsymbol{w}_{nt}\| \leq 1$ and

$$\Pr\{\boldsymbol{w}_{nt} = \boldsymbol{u}_{nt}, \quad t = 1, \ldots, n\} \longrightarrow 1 \tag{2.12}$$

as $n \to \infty$.

Now we shall verify that $x_{nt} = \boldsymbol{w}'_{nt}\boldsymbol{v}_t$ satisfy the conditions of Dvoretzky's theorem. We have

$$\mathcal{E}(\boldsymbol{w}'_{nt}\boldsymbol{v}_t|\mathcal{F}_{t-1}) = \boldsymbol{w}'_{nt}\mathcal{E}(\boldsymbol{v}_t|\mathcal{F}_{t-1}) = \mathbf{0} \quad \text{a.s.}, \tag{2.13}$$

$$\sum_{t=1}^{n} \mathcal{E}\left[(\boldsymbol{w}'_{nt}\boldsymbol{v}_t)^2|\mathcal{F}_{t-1}\right] = \sum_{t=1}^{n} \boldsymbol{w}'_{nt}\boldsymbol{\Sigma}_t\boldsymbol{w}_{nt} \xrightarrow{\text{p}} \text{tr } \boldsymbol{\Sigma D} \tag{2.14}$$

by (2.4). The third condition for $\{\boldsymbol{w}_{nt}\}$ to satisfy is

$$A_n(\delta) = \sum_{t=1}^{n} \mathcal{E}\left\{(\boldsymbol{w}'_{nt}\boldsymbol{v}_{nt})^2 I\left[(\boldsymbol{w}'_{nt}\boldsymbol{v}_{nt})^2 > \delta\right]|\mathcal{F}_{t-1}\right\} \xrightarrow{\text{p}} 0 \ \forall\delta > 0; \tag{2.15}$$

that is, given $\delta > 0$, $\varepsilon > 0$, and $\gamma > 0$, there exists $n(\varepsilon, \gamma)$ such that for $n > n\ (\varepsilon, \gamma)$

$$\Pr\{A_n(\delta) < \varepsilon\} > 1 - \gamma. \tag{2.16}$$

We have

$$\begin{aligned} A_n(\delta) &= \sum_{t=1}^{n} \boldsymbol{w}'_{nt}\boldsymbol{w}_{nt}\mathcal{E}\left\{\left(\frac{\boldsymbol{w}'_{nt}}{\|\boldsymbol{w}_{nt}\|}\boldsymbol{v}_t\right)^2 I\left[\left(\frac{\boldsymbol{w}'_{nt}}{\|\boldsymbol{w}_{nt}\|}\boldsymbol{v}_t\right)^2 > \frac{\delta}{\|\boldsymbol{w}_{nt}\|^2}\right] \middle|\mathcal{F}_{t-1}\right\} \\ &\leq \sum_{t=1}^{n} \boldsymbol{w}'_{nt}\boldsymbol{w}_{nt}\mathcal{E}\left\{\boldsymbol{v}'_t\boldsymbol{v}_t I\left[\boldsymbol{v}'_t\boldsymbol{v}_t > \frac{\delta}{\|\boldsymbol{w}_{nt}\|^2}\right] \middle|\mathcal{F}_{t-1}\right\}. \end{aligned} \tag{2.17}$$

Given $\varepsilon^* > 0$ and $\gamma^* > 0$ there exists $n^*(\varepsilon^*, \gamma^*)$ such that for $n > n^*(\varepsilon^*, \gamma^*)$

$$\Pr\left\{\|\boldsymbol{w}_{nt}\|^2 \le \varepsilon^*, t = 1, \ldots, n\right\} > 1 - \gamma^*. \tag{2.18}$$

Hence

$$\Pr\left\{A_n(\delta) \le \sum_{t=1}^{n} \boldsymbol{w}'_{nt}\boldsymbol{w}_{nt}\mathcal{E}\left[\boldsymbol{v}'_t\boldsymbol{v}_t I\left(\boldsymbol{v}'_t\boldsymbol{v}_t > \frac{\delta}{\varepsilon^*}\right)\Big|\mathcal{F}_{t-1}\right]\right\} \ge 1 - \gamma^*. \tag{2.19}$$

Since

$$\begin{aligned}
&\sum_{t=1}^{n} \boldsymbol{w}'_{nt}\boldsymbol{w}_{nt}\mathcal{E}\left\{\boldsymbol{v}'_t\boldsymbol{v}_t I\left(\boldsymbol{v}'_t\boldsymbol{v}_t > \frac{\delta}{\varepsilon^*}\right)\Big|\mathcal{F}_{t-1}\right\} \\
&\qquad \le \sum_{t=1}^{n} \boldsymbol{x}'_{nt}\boldsymbol{x}_{nt} \sup_s \mathcal{E}\left\{\boldsymbol{v}'_s\boldsymbol{v}_s I\left(\boldsymbol{v}'_s\boldsymbol{v}_s > \frac{\delta}{\varepsilon^*}\right)\Big|\mathcal{F}_{s-1}\right\} \\
&\qquad = B_n\left(\frac{\delta}{\varepsilon^*}\right),
\end{aligned} \tag{2.20}$$

say. That is,

$$\Pr\left\{A_n(\delta) \le B_n\left(\frac{\delta}{\varepsilon^*}\right)\right\} \ge 1 - \gamma^* \tag{2.21}$$

if $n > n^*\ (\varepsilon^*, \gamma^*)$. Let

$$C(d) = \sup_{s=1,2,\ldots} \mathcal{E}\left[\boldsymbol{v}'_s\boldsymbol{v}_s I(\boldsymbol{v}'_s\boldsymbol{v}_s > d)|\mathcal{F}_{s-1}\right]. \tag{2.22}$$

Condition (2.5) is that given $\epsilon > 0$, $\bar{\bar{\gamma}} > 0$ there exists a $d(\epsilon, \bar{\bar{\gamma}})$ such that for $d > d(\epsilon, \bar{\bar{\gamma}})$

$$\Pr\left\{C(d) \le \epsilon\right\} \ge 1 - \bar{\bar{\gamma}}. \tag{2.23}$$

Condition (2.2) implies that given $a > 0$, $\bar{\gamma} > 0$ there exists $\bar{n}(a, \bar{\gamma})$ such that

$$\Pr\left\{\sum_{t=1}^{n} \boldsymbol{x}'_{nt}\boldsymbol{x}_{nt} \le \operatorname{tr} \boldsymbol{D} + a\right\} \ge 1 - \bar{\gamma}. \tag{2.24}$$

Hence

$$\Pr\left\{B_n\left(\frac{\delta}{\epsilon^*}\right) < \epsilon\right\} \le 1 - \bar{\gamma} - \bar{\bar{\gamma}} \tag{2.25}$$

if $(\operatorname{tr} \boldsymbol{D} + a)\epsilon \le \varepsilon$, $\delta/\varepsilon^* \ge d(\epsilon, \gamma)$, and $n \ge \bar{n}(a, \bar{\gamma})$. Then (2.16) holds if $\gamma^* + \bar{\gamma} + \bar{\bar{\gamma}} \le \gamma$, $(\operatorname{tr} \boldsymbol{D} + a)\epsilon \le \varepsilon$, $\varepsilon^* \le \delta/d(\epsilon, \bar{\bar{\gamma}})$, and $n > \max\left[n^*(\varepsilon^*, \gamma^*), \bar{n}(a, \bar{\gamma})\right]$. The theorem follows from the theorem in the introduction [Dvoretzky (1972)]. [See, also, Corollary 3.1 of Hall and Heyde (1980) or Theorem 2, Section 9.5, of Chow and Teicher (1988).] ∎

Theorem 2. Let $\{\boldsymbol{v}_t\}$ be a sequence of random vectors and let $\{\mathcal{F}_t\}$ be an increasing sequence of σ-fields such that $\boldsymbol{v}_t$ is $\mathcal{F}_t$-measurable, $\mathcal{E}(\boldsymbol{v}_t|\mathcal{F}_{t-1}) = \boldsymbol{0}$ a.s., $\mathcal{E}(\boldsymbol{v}_t\boldsymbol{v}_t'|\mathcal{F}_{t-1}) = \boldsymbol{\Sigma}_t$ a.s.,

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{\Sigma}_t \xrightarrow{\text{p}} \boldsymbol{\Sigma}, \tag{2.26}$$

constant, and

$$\frac{1}{n}\sum_{t=1}^{n} \mathcal{E}\left[\boldsymbol{v}_t'\boldsymbol{v}_t I(\boldsymbol{v}_t'\boldsymbol{v}_t > n\varepsilon)|\mathcal{F}_{t-1}\right] \xrightarrow{\text{p}} 0. \tag{2.27}$$

Then

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{v}_t\boldsymbol{v}_t' \xrightarrow{\text{p}} \boldsymbol{\Sigma}. \tag{2.28}$$

Proof. If $\boldsymbol{v}_t$ is scalar, the proof follows from Theorem 2.23 of Hall and Heyde (1980) as indicated by Chan and Wei (1987). The theorem is then verified by taking arbitrary linear combinations of $\boldsymbol{v}_t$. ∎

Theorem 3. For n observations on the model (2.1) define

$$\hat{\boldsymbol{B}}_n = \sum_{t=1}^{n} \boldsymbol{y}_t\boldsymbol{z}_t' \left(\sum_{t=1}^{n} \boldsymbol{z}_t\boldsymbol{z}_t'\right)^{-1}, \tag{2.29}$$

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_n &= \frac{1}{n}\sum_{t=1}^{n} (\boldsymbol{y}_t - \hat{\boldsymbol{B}}_n\boldsymbol{z}_t)(\boldsymbol{y}_t - \hat{\boldsymbol{B}}_n\boldsymbol{z}_t)' \\ &= \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{v}_t\boldsymbol{v}_t' - \frac{1}{n}(\hat{\boldsymbol{B}}_n - \boldsymbol{B})\sum_{t=1}^{n} \boldsymbol{z}_t\boldsymbol{z}_t'(\hat{\boldsymbol{B}}_n - \boldsymbol{B})'. \end{aligned} \tag{2.30}$$

If the conditions of Theorem 1 hold with C nonsingular, then

$$\text{vec}\left[(\hat{B}_n - B)D_n\right] \xrightarrow{\mathcal{L}} N(0, C^{-1} \otimes \Sigma). \tag{2.31}$$

If, further, (2.26) holds, then

$$\hat{\Sigma}_n \xrightarrow{p} \Sigma. \tag{2.32}$$

Proof. The proof of (2.31) is a straightforward application of Theorem 1. The second term on the right-hand side of (2.30) is

$$\frac{1}{n}(\hat{B}_n - B)D_n^{-1}\left[D_n^{-1}\sum_{t=1}^{n} z_t z_t'(D_n')^{-1}\right]\left[(\hat{B}_n - B)D_n^{-1}\right]' \xrightarrow{p} 0 \tag{2.33}$$

by (2.2) and (2.31). ∎

The purpose of condition (2.3) is to assure asymptotic negligibility of $z_t v_t'$. What alternative conditions imply (2.3)?

Lemma 1. Let $\{z_t\}$ be a sequence of random vectors, and let $\{\mathcal{F}_t\}$ be an increasing sequence of σ-fields such that z_t is $\mathcal{F}_t$-measurable. Let D_n be $\mathcal{F}_0$-measurable such that $D_n^{-1} \to 0$ a.s., $D_n D_{n+1}^{-1} \xrightarrow{p} I$ a.s., and

$$D_n^{-1}\sum_{t=1}^{n} z_t z_t'(D_n')^{-1} \to C \quad \text{a.s.} \tag{2.34}$$

Then

$$\max_{t=1,\ldots,n} z_t'(D_n D_n')^{-1} z_t \to 0 \quad \text{a.s.} \tag{2.35}$$

Proof. From (2.34) we have

$$\begin{aligned} &D_{n+1}^{-1}\sum_{t=1}^{n+1} z_t z_t'(D_{n+1})^{-1} - D_n^{-1}\sum_{t=1}^{n} z_t z_t' D_n^{-1} \\ &= D_n^{-1} z_{n+1} z_{n+1}'(D_n')^{-1} + D_{n+1}^{-1}\sum_{t=1}^{n+1} z_t z_t'(D_{n+1}')^{-1} \\ &\quad - (D_n^{-1}D_{n+1})D_{n+1}^{-1}\sum_{t=1}^{n+1} z_t z_t'(D_{n+1}')^{-1}(D_{n+1}^{-1}D_{n+1})' \\ &\to 0 \quad \text{a.s.} \end{aligned} \tag{2.36}$$

That is, $\|\boldsymbol{D}_n^{-1}\boldsymbol{z}_{n+1}\|^2 \to 0$ a.s. This implies (2.35) by the proof of Lemma 2.6.1 in Anderson (1971). ∎

A special case of $\{\boldsymbol{z}_t\}$ is that of $\boldsymbol{z}_t$ nonstochastic; then (2.34) (which is the same as (2.2) when $\{\boldsymbol{z}_t\}$ is nonstochastic) implies (2.35) with the limits nonstochastic. In particular, if $\boldsymbol{D}_n$ is diagonal and the j-th diagonal element of $\boldsymbol{D}_n$ is the square root of the sum of squares of the j-th elements of the $\boldsymbol{z}_t$'s, then $\boldsymbol{D}_n^{-1}\sum_{t=1}^n \boldsymbol{z}_t\boldsymbol{z}_t'(\boldsymbol{D}_n')^{-1}$ is the correlation matrix of $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$. The theorem in this case is a relaxation of Theorems 2.6.1 and 2.6.2 of Anderson (1971).

Theorem 4. Let $\{\boldsymbol{z}_t\}$ be a sequence of random vectors, and let $\{\mathcal{F}_t\}$ be an increasing sequence of σ-fields such that $\boldsymbol{z}_t$ is $\mathcal{F}_t$-measurable and

$$\sum_{t=1}^n \mathcal{E}\left\{\boldsymbol{z}_t'(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1}\boldsymbol{z}_t I\left[\boldsymbol{z}_t'(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1}\boldsymbol{z}_t > \varepsilon\right]\middle|\mathcal{F}_{t-1}\right\} \xrightarrow{\mathrm{p}} 0. \tag{2.37}$$

Then (2.3) holds.

Proof. We use Lemma 3.5 of Dvoretzky (1972): If $\{\mathcal{F}_t\}$ is an increasing sequence of σ-fields and $A_t \in \mathcal{F}_t$, then for every $\eta > 0$

$$\Pr\left\{\bigcup_{t=1}^n A_t \middle| \mathcal{F}_0\right\} \leq \eta + \Pr\left\{\sum_{t=1}^n P(A_t|\mathcal{F}_{t-1}) > \eta \middle| \mathcal{F}_0\right\}. \tag{2.38}$$

For every $\varepsilon > 0$, $\eta > 0$

$$\begin{aligned}\Pr\left\{\max_{t=1,\ldots,n} \boldsymbol{z}_t'(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1}\boldsymbol{z}_t > \varepsilon\middle|\mathcal{F}_0\right\} &= \Pr\left\{\bigcup_{t=1}^n\left[\boldsymbol{z}_t'(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1}\boldsymbol{z}_t > \varepsilon\middle|\mathcal{F}_0\right]\right\} \\ &\leq \eta + \Pr\left\{\sum_{t=1}^n \Pr\left(\boldsymbol{z}_t'(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1}\boldsymbol{z}_t > \varepsilon\middle|\mathcal{F}_{t-1}\right) > \eta\middle|\mathcal{F}_0\right\} \\ &\leq \eta + \Pr\left\{\frac{1}{n}\sum_{t=1}^n \mathcal{E}\left[\boldsymbol{z}_t'(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1}\boldsymbol{z}_t I\left[\boldsymbol{z}_t'(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1}\boldsymbol{z}_t > \varepsilon\middle|\mathcal{F}_{t-1}\right] > \eta\middle|\mathcal{F}_0\right\}\end{aligned} \tag{2.39}$$

by a form of Tchebycheff's inequality. By (2.37) the right-hand side of (2.39) converges to 0. Since η is arbitrary, (2.3) holds. ∎

Corollary 1. Let $\{\boldsymbol{z}_t, \boldsymbol{v}_t\}$, $t = 1, 2, \ldots$, be a sequence of pairs of random vectors, and let $\{\mathcal{F}_t\}$ be an increasing sequence of σ-fields such that $\boldsymbol{z}_t$ is $\mathcal{F}_{t-1}$-measurable and $\boldsymbol{v}_t$ is $\mathcal{F}_t$-measurable. Let $\boldsymbol{D}_n$ be $\mathcal{F}_0$-measurable such that (2.2) and (2.37) hold. Suppose that $\mathcal{E}(\boldsymbol{v}_t|\mathcal{F}_{t-1}) = 0$ a.s., $\mathcal{E}(\boldsymbol{v}_t\boldsymbol{v}_t'|\mathcal{F}_{t-1}) = \boldsymbol{\Sigma}_t$ a.s., and (2.4) and (2.5) hold. Then (2.6) holds.

The condition (2.4) determines the limiting covariance matrix of $\boldsymbol{D}_n^{-1}\sum_{t=1}^n \boldsymbol{z}_t\boldsymbol{v}_t'$.

Lemma 2. Let $\{\boldsymbol{z}_t, \boldsymbol{v}_t\}$ be a sequence of random vectors, and let $\{\mathcal{F}\}$ be an increasing sequence of σ-fields such that $\boldsymbol{z}_t$ is $\mathcal{F}_{t-1}$-measurable and $\boldsymbol{v}_t$ is $\mathcal{F}_t$-measurable such that $\mathcal{E}(\boldsymbol{v}_t|\mathcal{F}_{t-1}) = \boldsymbol{0}$ a.s., $\mathcal{E}(\boldsymbol{v}_t\boldsymbol{v}_t'|\mathcal{F}_{t-1}) = \boldsymbol{\Sigma}_t$ a.s., and $\boldsymbol{\Sigma}_t \to \boldsymbol{\Sigma}$ a.s., where $\boldsymbol{\Sigma}$ is a constant matrix. Suppose $\boldsymbol{D}_n$ is $\mathcal{F}_0$-measurable such that (2.2) holds. Then (2.4) and (2.26) hold. If, further, (2.3) and (2.5) hold, then (2.6) holds.

The homoscedastic case, $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}$, is included and also the case of $\boldsymbol{\Sigma}_t$ nonstochastic.

An important case of $\{\boldsymbol{z}_t\}$ is that in which $\boldsymbol{D}_n = \sqrt{n}\,\boldsymbol{I}$; then $\boldsymbol{D}_n^{-1}\sum_{t=1}^n \boldsymbol{z}_t\boldsymbol{z}_t'(\boldsymbol{D}_n')^{-1} = (1/n)\sum_{t=1}^n \boldsymbol{z}_t\boldsymbol{z}_t'$; that is, this matrix is simply the sample covariance matrix for known mean $\boldsymbol{0}$.

Corollary 2. Let $\{\boldsymbol{z}_t, \boldsymbol{v}_t\}$ be a sequence of pairs of random vectors and let $\{\mathcal{F}_t\}$ be an increasing sequence of σ-fields such that $\boldsymbol{z}_t$ is $\mathcal{F}_{t-1}$-measurable and $\boldsymbol{v}_t$ is $\mathcal{F}_t$-measurable. Suppose

$$\frac{1}{n}\sum_{t=1}^n \boldsymbol{z}_t\boldsymbol{z}_t' \xrightarrow{\mathrm{p}} \boldsymbol{M}, \tag{2.40}$$

a constant matrix,

$$\frac{1}{n}\max_{t=1,\ldots,n} \boldsymbol{z}_t'\boldsymbol{z}_t \xrightarrow{\mathrm{p}} 0, \tag{2.41}$$

$\mathcal{E}(\boldsymbol{v}_t|\mathcal{F}_{t-1}) = \boldsymbol{0}$ a.s., $\mathcal{E}(\boldsymbol{v}_t\boldsymbol{v}_t'|\mathcal{F}_{t-1}) = \boldsymbol{\Sigma}_t$ a.s.,

$$\frac{1}{n}\sum_{t=1}^n (\boldsymbol{\Sigma}_t \otimes \boldsymbol{z}_t\boldsymbol{z}_t') \xrightarrow{\mathrm{p}} \boldsymbol{\Sigma} \otimes \boldsymbol{M}, \tag{2.42}$$

and (2.5) holds. Then

$$\frac{1}{\sqrt{n}}\operatorname{vec}\left(\sum_{t=1}^n \boldsymbol{z}_t\boldsymbol{v}_t'\right) \xrightarrow{\mathcal{L}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{M}); \tag{2.43}$$

if, further, M is nonsingular, then

$$\sqrt{n}\ \mathrm{vec}(\hat{B}_n - B) \xrightarrow{\mathcal{L}} N(\mathbf{0}, M^{-1} \otimes \Sigma); \tag{2.44}$$

and if, further, (2.26) holds, then (2.32) holds.

Condition (2.40) is equivalently $(1/n)\sum_{t=1}^{n} \mathrm{vec}\ z_t z_t' \xrightarrow{\mathrm{P}} \mathrm{vec}\ M$; (2.26) is equivalently $(1/n)\,\mathrm{vec}\ \Sigma_t \xrightarrow{\mathrm{P}} \mathrm{vec}\ \Sigma$; and (2.42) is equivalently

$$\frac{1}{n}\sum_{t=1}^{n} \mathrm{vec}\ \Sigma_t(\mathrm{vec}\ z_t z_t')' - \frac{1}{n}\sum_{t=1}^{n} \mathrm{vec}\ \Sigma_t \left(\frac{1}{n}\sum_{t=1}^{n} \mathrm{vec}\ z_t z_t'\right)' \xrightarrow{\mathrm{P}} \mathbf{0}. \tag{2.45}$$

The condition (2.45) is that vec Σ_t and vec $z_t z_t'$ are asymptotically uncorrelated over t. Even if the Σ_t's are nonstochastic and the z_t are exogenous this condition is needed to obtain $\Sigma \otimes M$ as the covariance matrix of $(1/\sqrt{n})\,\mathrm{vec}\ \sum_{t=1}^{n} z_t v_t'$.

3. Robustness in Autoregression.

We now consider the autoregressive model.

$$x_t = Bx_{t-1} + v_t, \quad t = 1, 2, \ldots\ . \tag{3.1}$$

The form of (3.1) is (2.1) with z_t replaced by x_{t-1}. We shall show that the least squares estimator of B based on $x_0, \ldots, x_n$ has the asymptotic normal distribution of the least squares estimator in the regression case. In order to show the analogies to (2.2) and (2.3) we prove the following lemmas.

Lemma 3. If the characteristic roots of B are less than 1 in absolute value and if $\max_{t=1,\ldots,n} v_t'v_t/n \xrightarrow{\mathrm{P}} 0$, then for $x_1, x_2, \ldots$ generated by (3.1)

$$\frac{1}{n}\max_{t=1,\ldots,n} x_{t-1}'x_{t-1} \xrightarrow{\mathrm{P}} 0. \tag{3.2}$$

Proof. Since $x_0'x_0/n \xrightarrow{\mathrm{P}} 0$ and the roots of B are less than 1 in absolute value, $x_0'(B')^{t-1}B^{t-1}x_0/n \xrightarrow{\mathrm{P}} 0$ and we need only consider

$$x_{t-1}^* = \sum_{s=0}^{t-2} B^s v_{t-1-s}. \tag{3.3}$$

Then

$$\begin{aligned}
(3.4) \qquad \boldsymbol{x}^{*\prime}_{t-1}\boldsymbol{x}^{*}_{t-1} &= \sum_{r,s=0}^{t-2} \boldsymbol{v}'_{t-r-1}(\boldsymbol{B}')^r \boldsymbol{B}^s \boldsymbol{v}_{t-s-1} \\
&\leq \sum_{r,s=0}^{t-2} \left|\boldsymbol{v}'_{t-r-1}(\boldsymbol{B}')^r \boldsymbol{B}^s \boldsymbol{v}_{t-s-1}\right| \\
&\leq \sum_{r,s=0}^{t-2} \lambda^{r+s} q r^{p-1} s^{p-1} \left(\|\boldsymbol{v}_{t-r-1}\|^2 + \|\boldsymbol{v}_{t-s-1}\|^2\right),
\end{aligned}$$

where λ is the largest absolute value of the characteristic roots of $\boldsymbol{B}$ and q is a suitable constant. (See Lemma 7 in the appendix.) Then

$$(3.5) \qquad \frac{1}{n} \max_{t=1,\ldots,n} \|\boldsymbol{x}^{*}_{t-1}\|^2 \leq \frac{2q}{n} \max_{t=1,\ldots,n} \|\boldsymbol{v}_t\|^2 \left(\sum_{s=0}^{n-2} \lambda^s s^{p-1}\right)^2 .$$

Since the sum in (3.5) is bounded as $n \to \infty$, (3.2) follows. ∎

Lemma 4. Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ be generated by (3.1) with and $\mathcal{E}\boldsymbol{x}_0\boldsymbol{x}'_0 = \boldsymbol{\Sigma}_0$. Let $\{\mathcal{F}_t\}$ be an increasing sequence of σ-fields such that $\boldsymbol{x}_t$ and $\boldsymbol{v}_t$ are $\mathcal{F}_t$-measurable. Suppose the characteristic roots of $\boldsymbol{B}$ are less than 1 in absolute value, $\mathcal{E}(\boldsymbol{v}_t|\mathcal{F}_{t-1}) = \boldsymbol{0}$ a.s., $\mathcal{E}(\boldsymbol{v}_t\boldsymbol{v}'_t|\mathcal{F}_{t-1}) = \boldsymbol{\Sigma}_t$ a.s., (2.26) holds with $\boldsymbol{\Sigma}$ constant, and (2.27) holds. Define

$$(3.6) \qquad \boldsymbol{\Gamma} = \sum_{s=0}^{\infty} \boldsymbol{B}^s \boldsymbol{\Sigma} (\boldsymbol{B}')^s .$$

Then (2.28) holds,

$$(3.7) \qquad \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{v}_t \boldsymbol{x}'_{t-1} \xrightarrow{\text{P}} \boldsymbol{0},$$

$$(3.8) \qquad \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_{t-1} \boldsymbol{x}'_{t-1} \xrightarrow{\text{P}} \boldsymbol{\Gamma}.$$

Proof. From (3.1) we have

$$(3.9) \qquad \boldsymbol{x}_{t-1} = \sum_{t=0}^{t-2} \boldsymbol{B}^s \boldsymbol{v}_{t-1-s} + \boldsymbol{B}^{t-1}\boldsymbol{x}_0 .$$

For some $\theta > 0$ define $\boldsymbol{x}_{n0} = \boldsymbol{x}_0$.

$$\boldsymbol{v}_{nt} = \boldsymbol{v}_t I\left[\operatorname{tr}\sum_{s=1}^{t}\boldsymbol{\Sigma}_s \leq n(1+\theta)\operatorname{tr}\,\boldsymbol{\Sigma}_s\right]. \tag{3.10}$$

$$\boldsymbol{x}_{n,t-1} = \sum_{s=0}^{t-2}\boldsymbol{B}^s\boldsymbol{v}_{n,t-1-s} + \boldsymbol{B}^{t-1}\boldsymbol{x}_{n0}. \tag{3.11}$$

Then

$$\Pr\left\{\boldsymbol{v}_{nt} = \boldsymbol{v}_t, t = 1,\ldots,n\right\} \xrightarrow{\text{p}} 1. \tag{3.12}$$

$$\Pr\left\{\boldsymbol{x}_{n,t-1} = \boldsymbol{x}_{t-1}, t = 1,\ldots,n\right\} \xrightarrow{\text{p}} 1, \tag{3.13}$$

$$\Pr\left\{\boldsymbol{v}_{nt}\boldsymbol{x}'_{n,t-1} = \boldsymbol{v}_t\boldsymbol{x}'_{t-1}, t = 1,\ldots,n\right\} \xrightarrow{\text{p}} 1. \tag{3.14}$$

By construction $\mathcal{E}\|\boldsymbol{v}_{nt}\|^2 \leq n(1+\theta)\operatorname{tr}\,\boldsymbol{\Sigma}$ and $\mathcal{E}\|\boldsymbol{x}_{n,t-1}\|^2 < \infty$. Then

$$\begin{aligned}
\operatorname{tr}\,\mathcal{E}\left(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{v}_{nt}\boldsymbol{x}'_{n,t-1}\right)\left(\frac{1}{n}\sum_{s=1}^{n}\boldsymbol{x}_{n,s-1}\boldsymbol{v}'_{ns}\right) \tag{3.15}\\
&= \frac{1}{n^2}\mathcal{E}\operatorname{tr}\sum_{s,t=1}^{n}\boldsymbol{v}_{nt}\boldsymbol{x}'_{n,t-1}\boldsymbol{x}_{n,s-1}\boldsymbol{v}'_{ns}\\
&= \frac{1}{n^2}\operatorname{tr}\,\mathcal{E}\sum_{s,t=1}^{n}\boldsymbol{x}'_{n,t-1}\boldsymbol{x}_{n,s-1}\boldsymbol{v}'_{n,s}\boldsymbol{v}_{nt}\\
&= \frac{1}{n^2}\mathcal{E}\sum_{s,t=1}^{n}\boldsymbol{x}'_{n,t-1}\boldsymbol{x}_{n,s-1}\mathcal{E}\left(\boldsymbol{v}'_{ns}\boldsymbol{v}_{nt}|\mathcal{F}_{\max(s,t)-1}\right)\\
&= \frac{1}{n^2}\mathcal{E}\sum_{t=1}^{n}\boldsymbol{x}'_{n,t-1}\boldsymbol{x}_{n,t-1}\mathcal{E}\left(\boldsymbol{v}'_{nt}\boldsymbol{v}_{nt}|\mathcal{F}_{t-1}\right).
\end{aligned}$$

Since $\max_{t=1,\ldots,n}\|\boldsymbol{v}_t\|^2/n \xrightarrow{\text{p}} 0$ by Theorem 4, we have $\max_{t=1,\ldots,n}\mathcal{E}(\|\boldsymbol{v}_{nt}\|^2|\mathcal{F}_{t-1})/n \xrightarrow{\text{p}} 0$ by (3.6). Now consider for $2 \leq t \leq n-1$

$$\begin{aligned}
&\frac{1}{n}\mathcal{E}\sum_{t=1}^{n}\boldsymbol{x}_{n,t-1}\boldsymbol{x}'_{n,t-1} \tag{3.16}\\
&= \frac{1}{n}\mathcal{E}\left[\boldsymbol{x}_0\boldsymbol{x}'_0 + \sum_{t=2}^{n}\left(\sum_{r=0}^{t-2}\boldsymbol{B}^r\boldsymbol{v}_{n,t-r-1} + \boldsymbol{B}^{t-1}\boldsymbol{x}_0\right)\left(\sum_{s=0}^{t-2}\boldsymbol{B}^s\boldsymbol{v}_{n,t-s-1} + \boldsymbol{B}^{t-1}\boldsymbol{x}_0\right)'\right]\\
&= \frac{1}{n}\sum_{t=2}^{n}\sum_{s=0}^{t-2}\boldsymbol{B}^s\mathcal{E}\boldsymbol{v}_{n,t-s-1}\boldsymbol{v}'_{n,t-s-1}(\boldsymbol{B}')^s + \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{B}^{t-1}\boldsymbol{\Sigma}_0(\boldsymbol{B}')^{t-1}\\
&= \sum_{s=0}^{n-2}\boldsymbol{B}^s\frac{1}{n}\sum_{t=s+2}^{n}\mathcal{E}\boldsymbol{v}_{n,t-s-1}\boldsymbol{v}'_{n,t-s-1}(\boldsymbol{B}')^s + \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{B}^{t-1}\boldsymbol{\Sigma}_0(\boldsymbol{B}')^{t-1}.
\end{aligned}$$

The trace of the first term on the right-hand side of (3.6) is not greater than $(1+\theta)\,\mathrm{tr}\,\boldsymbol{\Gamma}$. Hence, (3.5) $\to 0$, and (3.7) is proved.

From (2.28) and (3.1) we have

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{v}_t\boldsymbol{v}_t' &= \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{x}_t\boldsymbol{x}_t' - \boldsymbol{x}_t\boldsymbol{x}_{t-1}'\boldsymbol{B}' - \boldsymbol{B}\boldsymbol{x}_t\boldsymbol{x}_{t-1}' + \boldsymbol{B}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\boldsymbol{B}') \\ &\xrightarrow{\mathrm{p}} \boldsymbol{\Sigma}.\end{aligned} \tag{3.17}$$

From (3.7) and (3.1) we have

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{v}_t\boldsymbol{x}_{t-1}' &= \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{x}_t\boldsymbol{x}_{t-1}' - \boldsymbol{B}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}') \\ &\xrightarrow{\mathrm{p}} \mathbf{0}.\end{aligned} \tag{3.18}$$

If we add to (3.17) the result of multiplying (3.18) on the right by $\boldsymbol{B}'$ and the transpose of that product, we obtain

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{v}_t\boldsymbol{v}_t' + \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{v}_t\boldsymbol{x}_{t-1}'\boldsymbol{B}' + \frac{1}{n}\boldsymbol{B}\sum_{t=1}^{n}\boldsymbol{x}_{t-1}\boldsymbol{v}_t' \\ = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t' - \boldsymbol{B}\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\boldsymbol{B}' \\ \xrightarrow{\mathrm{p}} \boldsymbol{\Sigma}.\end{aligned} \tag{3.19}$$

Furthermore, Lemma 3 implies

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t' - \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_{t-1}x_{t-1} = \frac{1}{n}\boldsymbol{x}_n\boldsymbol{x}_n' - \frac{1}{n}\boldsymbol{x}_0\boldsymbol{x}_0' \xrightarrow{\mathrm{p}} \mathbf{0} \tag{3.20}$$

Then (3.19) is equivalent to

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t' - \boldsymbol{B}\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t'\boldsymbol{B}' \xrightarrow{\mathrm{p}} \boldsymbol{\Sigma}, \tag{3.21}$$

which implies

$$\boldsymbol{\Gamma} = \mathrm{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t' = \mathrm{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'. \tag{3.22}$$

See Problem 27 of Chapter 5 of Anderson (1971). Then (3.4) follows. ∎

Theorem 5. Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ be generated by (3.1), where $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$ is a sequence of random vectors and $\mathcal{E}\boldsymbol{x}_0\boldsymbol{x}_0' = \boldsymbol{\Sigma}_0$. Let $\{\mathcal{F}_t\}$ be an increasing sequence of σ-fields such that $\boldsymbol{x}_t$ and $\boldsymbol{v}_t$ are $\mathcal{F}_t$-measurable. Suppose that the characteristic roots of $\boldsymbol{B}$ are less than 1 in absolute value, $\mathcal{E}(\boldsymbol{v}_t|\mathcal{F}_{t-1}) = \boldsymbol{0}$ a.s. $\mathcal{E}(\boldsymbol{v}_t\boldsymbol{v}_t'|\mathcal{F}_{t-1}) = \boldsymbol{\Sigma}_t$ a.s., (2.26) holds with $\boldsymbol{\Sigma}$ constant, and (2.5) holds. Furthermore, suppose

$$\frac{1}{n}\sum_{t=\max(r,s)+2}^{n}\left(\boldsymbol{\Sigma}_t \otimes \boldsymbol{v}_{t-1-r}\boldsymbol{v}_{t-1-s}'\right) \xrightarrow{\text{p}} \delta_{rs}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}), \tag{3.23}$$

where $\delta_{ss} = 1$ and $\delta_{rs} = 0$ for $r \neq s$. Then

$$\frac{1}{\sqrt{n}}\,\text{vec}\left(\sum_{t=1}^{n}\boldsymbol{x}_{t-1}\boldsymbol{v}_t'\right) \xrightarrow{\mathcal{L}} N(\boldsymbol{0}, \boldsymbol{\Sigma}\otimes\Gamma). \tag{3.24}$$

Proof. In Corollary 2 we take $\boldsymbol{z}_t = \boldsymbol{x}_{t-1}$. We want to verify (2.40), (2.41), and (2.42); (2.5) is assumed. Since (2.5) implies (2.27), Lemma 4 includes (3.8), which is equivalent to (2.40).

We have

$$\begin{aligned}(3.25)\quad &\frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{\Sigma}_t\otimes\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}')\\ &= \frac{1}{n}\sum_{t=1}^{n}\left[\boldsymbol{\Sigma}_t\otimes\left(\sum_{r=0}^{t-2}\boldsymbol{B}^r\boldsymbol{v}_{t-r-1}+\boldsymbol{B}^{t-1}\boldsymbol{x}_0\right)\left(\sum_{s=0}^{t-2}\boldsymbol{B}^s\boldsymbol{v}_{t-s-1}+\boldsymbol{B}^{t-1}\boldsymbol{x}_0\right)'\right]\end{aligned}$$

If we define $\boldsymbol{v}_0 = \boldsymbol{v}_{-1} = \cdots = \boldsymbol{0}$, we can write

$$\begin{aligned}(3.26)\quad \sum_{s=0}^{t-2}\boldsymbol{B}^s\boldsymbol{v}_{t-s-1}+\boldsymbol{B}^{t-1}\boldsymbol{x}_0 &= \sum_{s=0}^{\infty}\boldsymbol{B}^s\boldsymbol{v}_{t-s-1}+\boldsymbol{B}^{t-1}\boldsymbol{x}_0\\ &= \sum_{s=0}^{k}\boldsymbol{B}^s\boldsymbol{v}_{t-s-1}+\sum_{s=k+1}^{\infty}\boldsymbol{B}^s\boldsymbol{v}_{t-s-1}+\boldsymbol{B}^{t-1}\boldsymbol{x}_0.\end{aligned}$$

For $t \geq p+1$

$$\|\boldsymbol{B}^{t-1}\boldsymbol{x}_0\| \leq 2\lambda^{2(t-1)}q(t-1)^{p-1}\|\boldsymbol{x}_0\|^2. \tag{3.27}$$

Hence

$$\frac{1}{n}\sum_{t=1}^{n}\left[\boldsymbol{\Sigma}_t \otimes \boldsymbol{B}^{t-1}\boldsymbol{x}_0\boldsymbol{x}_0'(\boldsymbol{B}')^{t-1}\right] \xrightarrow{\text{p}} \mathbf{0}. \tag{3.28}$$

(See Lemma 8 in the Appendix.)

Consider the positive semidefinite matrix

$$\frac{1}{n}\sum_{t=1}^{n}\left[\boldsymbol{\Sigma}_t \otimes \sum_{r,s=k+1}^{\infty} \boldsymbol{B}^r \boldsymbol{v}_{t-r-1}\boldsymbol{v}_{t-s-1}'(\boldsymbol{B}')^s\right]. \tag{3.29}$$

We shall show that with arbitrarily high probability the trace of (3.28) is arbitrarily small if k is large enough. That will follow by showing the same property of

$$\frac{1}{n}\sum_{t=1}^{n}\left[\boldsymbol{\Sigma}_{nt} \otimes \sum_{r,s=k+1}^{\infty} \boldsymbol{B}^r \boldsymbol{v}_{n,t-r-1}\boldsymbol{v}_{t-s-1}'(\boldsymbol{B}')^s\right], \tag{3.30}$$

where $\boldsymbol{\Sigma}_{nt} = \mathcal{E}(\boldsymbol{v}_{nt}\boldsymbol{v}_{nt}'|\mathcal{F}_{t-1})$. The expected value of the trace of the second matrix in (3.30) is

$$\begin{aligned}
(3.31)\quad \mathcal{E}\sum_{r,s=k+1}^{\infty} &\operatorname{tr}\, \boldsymbol{B}^r \boldsymbol{v}_{n,t-r-1}\boldsymbol{v}_{n,t-s-1}'(\boldsymbol{B})^s \\
&= \mathcal{E}\sum_{s=k+1}^{\infty} \boldsymbol{v}_{n,t-s-1}'(\boldsymbol{B}')^s\boldsymbol{B}^r\boldsymbol{v}_{n,t-r-1} \\
&\le \sum_{s=k+1}^{\infty} \lambda^{2s}q^* s^{2p}\mathcal{E}\boldsymbol{v}_{n,t-s-1}'\boldsymbol{v}_{n,t-s-1} \\
&= q^*\sum_{s=k+1}^{\infty}\lambda^{2s}s^{2p}\mathcal{E}\Big\{\mathcal{E}\big[\boldsymbol{v}_{n,t-s-1}'\boldsymbol{v}_{n,t-s-1}I(\boldsymbol{v}_{n,t-s-1}'\boldsymbol{v}_{n,t-s-1}\le a)|\mathcal{F}_{t-s-2}\big] \\
&\qquad + \mathcal{E}\big[\boldsymbol{v}_{n,t-s-1}'\boldsymbol{v}_{n,t-s-1}I(\boldsymbol{v}_{n,t-s-1}'\boldsymbol{v}_{n,t-s-1}> a)\big|\mathcal{F}_{t-s-2}\big]\Big\} \\
&\le q^*\sum_{s=k+1}^{\infty}\lambda^{2s}s^{2p}\left\{a + \mathcal{E}\sup_{t=1,2,\dots}\mathcal{E}\big[\boldsymbol{v}_{nt}'\boldsymbol{v}_{nt}I(\boldsymbol{v}_{nt}'\boldsymbol{v}_{nt}>a)|\mathcal{F}_{t-1}\big]\right\}.
\end{aligned}$$

Since $\sum_{s=k+1}^{\infty}\lambda^{2s}s^{2p}$ converges, the second part of the right–hand side of (3.31) can be made arbitrarily small by taking a large enough; the first term can be made arbitrarily small by making k sufficiently large. Thus (3.31) is arbitrarily small, and by Tchebycheff's inequality the second matrix in (3.30) is arbitrarily small with arbitrarily high probability.

Now

$$(3.32)\qquad \frac{1}{n}\sum_{t=1}^{n}\left[\boldsymbol{\Sigma}_t \otimes \sum_{r,s=0}^{k} \boldsymbol{B}^r \boldsymbol{v}_{t-r-1}\boldsymbol{v}'_{t-s-1}(\boldsymbol{B}')^s\right] \xrightarrow{\text{P}} \boldsymbol{\Sigma}\otimes\sum_{s=0}^{k}\boldsymbol{B}^s\boldsymbol{\Sigma}(\boldsymbol{B}')^s.$$

If the right-hand side of (3.26) is written as $\boldsymbol{a}_t + \boldsymbol{b}_t + \boldsymbol{c}_t$, we have shown above that

$$(3.33)\qquad \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\Sigma}_t\otimes\boldsymbol{b}_t\boldsymbol{b}'_t \xrightarrow{\text{P}} 0$$

and that

$$(3.34)\qquad \operatorname{tr}\frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{\Sigma}_t\otimes\boldsymbol{c}_t\boldsymbol{c}'_t)$$

can be made to converge in probability as $n\to\infty$ to an arbitrarily small quantity. It follows from the Cauchy-Schwarz inequality that

$$(3.35)\qquad \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{\Sigma}_t\otimes\boldsymbol{a}_t\boldsymbol{c}'_t) \xrightarrow{\text{P}} \mathbf{0},$$

$$(3.36)\qquad \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{\Sigma}_t\otimes\boldsymbol{b}_t\boldsymbol{c}'_t) \xrightarrow{\text{P}} \mathbf{0},$$

and that

$$(3.37)\qquad \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{\Sigma}_t\otimes\boldsymbol{a}_t\boldsymbol{b}'_t)$$

can be made to converge in probability to an arbitrarily small quantity. Hence,

$$(3.38)\qquad \frac{1}{n}\sum_{t=1}^{n}[\boldsymbol{\Sigma}_t\otimes\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}] \xrightarrow{\text{P}} \boldsymbol{\Sigma}\otimes\boldsymbol{\Gamma}.$$

Hence, by Corollary 2 (3.24) follows. ∎

The least squares estimator of $\boldsymbol{B}$ is

$$(3.39)\qquad \hat{\boldsymbol{B}}_n = \sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}'_{t-1}\left(\sum_{t=1}^{n}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\right)^{-1},$$

and the estimator of $\boldsymbol{\Sigma}$ is

$$\hat{\boldsymbol{\Sigma}}_n = \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{x}_t - \hat{\boldsymbol{B}}_n\boldsymbol{x}_{t-1})(\boldsymbol{x}_t - \hat{\boldsymbol{B}}_n\boldsymbol{x}_{t-1})' \tag{3.40}$$
$$= \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{v}_t\boldsymbol{v}_t' - (\hat{\boldsymbol{B}}_n - \boldsymbol{B})\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'(\hat{\boldsymbol{B}}_n - \boldsymbol{B})'.$$

Corollary 3. Suppose the conditions of Theorem 5 hold and $\boldsymbol{\Gamma}$ is nonsingular. Then

$$\sqrt{n}\,\mathrm{vec}(\hat{\boldsymbol{B}}_n - \boldsymbol{B}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}^{-1} \otimes \boldsymbol{\Sigma}), \tag{3.41}$$

and (2.32) holds.

The conditions (3.23) in autoregression replace condition (2.4) in regression; they imply (3.38) which is the analog of (2.4). The limit (3.38) is that vec $\boldsymbol{\Sigma}_t$ and vec $\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'$ are asymptotically uncorrelated. The condition holds identically in $\boldsymbol{B}$; the conditions (3.23) are independent of $\boldsymbol{B}$.

Corollary 4. Under the conditions of Theorem 5 with (2.26) and (3.23) replaced by $\boldsymbol{\Sigma}_t \to \boldsymbol{\Sigma}$ a.s., (3.24) holds. If $\boldsymbol{\Gamma}$ is nonsingular, (3.41) and (2.32) hold.

Proof. The condition $\boldsymbol{\Sigma}_t \to \boldsymbol{\Sigma}$ a.s., where $\boldsymbol{\Sigma}$ is constant, implies (2.26) and (3.23). ∎

A higher order autoregressive process can be reduced to the first–order process. Suppose $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ satisfy

$$\boldsymbol{X}_t = \boldsymbol{B}_1\boldsymbol{X}_{t-1} + \cdots + \boldsymbol{B}_p\boldsymbol{X}_{t-p} + \boldsymbol{V}_t, t = 1, 2, \ldots \, . \tag{3.42}$$

Define

$$\boldsymbol{x}_t = \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{X}_{t-1} \\ \vdots \\ \boldsymbol{X}_{t-p+1} \end{bmatrix}, \boldsymbol{v}_t = \begin{bmatrix} \boldsymbol{V}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \tag{3.43}$$

$$B = \begin{bmatrix} B_1 & B_2 & B_3 & \cdots & B_p \\ I & 0 & 0 & \cdots & 0 \\ 0 & I & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \Sigma_t = \begin{bmatrix} \Omega_t & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{3.44}$$

where $\mathcal{E}(V_t|\mathcal{F}_{t-1}) = 0$ a.s., $\mathcal{E}(V_t V_t'|\mathcal{F}_{t-1}) = \Omega_t$ a.s., and $\{\mathcal{F}_t\}$ is an increasing σ–field such that X_t and V_t are $\mathcal{F}_t$–measurable. Then $\{x_t\}$ satisfies (3.1).

Theorem 6. Let

$$\mathcal{E} \begin{bmatrix} X_0 \\ X_{-1} \\ \vdots \\ X_{-p+1} \end{bmatrix} [X_0', X_{-1}', \ldots, X_{-p+1}'] = \Phi, \tag{3.45}$$

and let $X_1, X_2, \ldots$ be generated by (3.42). Let $\{\mathcal{F}_t\}$ be an increasing sequence of σ–fields such that X_t and V_t are $\mathcal{F}_t$–measurable. Suppose the roots of

$$|\lambda^p I - \lambda^{p-1} B_1 - \cdots - B_p| = 0 \tag{3.46}$$

are less than 1 in absolute value, $\mathcal{E}(V_t|\mathcal{F}_{t-1}) = 0$ a.s., $\mathcal{E}(V_t V_t'|\mathcal{F}_{t-1}) = \Omega_t$ a.s.,

$$\frac{1}{n}\sum_{t=1}^{n} \Omega_t \xrightarrow{p} \Omega, \tag{3.47}$$

which is nonsingular and constant, and (2.5) holds with v_t replaced by V_t. Define

$$(\hat{B}_{1n}, \hat{B}_{2n}, \ldots, \hat{B}_{pn}) = \sum_{t=1}^{n} X_t(X_{t-1}', X_{t-2}', \ldots, X_{t-p}') \tag{3.48}$$
$$\times \begin{bmatrix} \sum_{t=1}^n X_{t-1}X_{t-1}' & \sum_{t=1}^n X_{t-1}X_{t-2}' & \cdots & \sum_{t=1}^n X_{t-1}X_{t-p}' \\ \sum_{t=1}^n X_{t-2}X_{t-1}' & \sum_{t=1}^n X_{t-2}X_{t-2}' & \cdots & \sum_{t=1}^n X_{t-2}X_{t-p}' \\ \vdots & \vdots & & \vdots \\ \sum_{t=1}^n X_{t-p}X_{t-1}' & \sum_{t=1}^n X_{t-p}X_{t-2}' & \cdots & \sum_{t=1}^n X_{t-p}X_{t-p}' \end{bmatrix}^{-1},$$

$$\hat{\Omega}_n = \frac{1}{n}\sum_{t=1}^{n}(X_t - \hat{B}_{1n}X_{t-1} - \cdots - \hat{B}_{pn}X_{t-p})(X_t - \hat{B}_{1n}X_{t-1} - \cdots - \hat{B}_{pn}X_{t-p})'. \tag{3.49}$$

Then

$$\hat{\Omega}_n \xrightarrow{p} \Omega, \tag{3.50}$$

$$
(3.51) \qquad \frac{1}{n}\sum_{t=1}^{n}\begin{bmatrix} \boldsymbol{X}_{t-1} \\ \boldsymbol{X}_{t-2} \\ \vdots \\ \boldsymbol{X}_{t-p} \end{bmatrix}\left[\boldsymbol{X}'_{t-1}, \boldsymbol{X}'_{t-2}, \ldots, \boldsymbol{X}'_{p-p}\right] \xrightarrow{\text{p}} \sum_{s=0}^{\infty}\boldsymbol{B}^s\boldsymbol{\Sigma}(\boldsymbol{B}')^s = \boldsymbol{\Gamma},
$$

say, where

$$
(3.52) \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Omega} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix},
$$

and

$$
(3.53) \qquad \sqrt{n}\,\mathrm{vec}(\hat{\boldsymbol{B}}_{1n} - \boldsymbol{B}_1, \ldots, \hat{\boldsymbol{B}}_{pn} - \boldsymbol{B}_p) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{\Gamma}^{-1} \otimes \boldsymbol{\Omega}).
$$

Lemma 5. If $\boldsymbol{\Omega}$ is nonsingular, $\boldsymbol{\Gamma}$ is nonsingular.

Proof. The proof is a vector generalization of the proof of Lemma 5.5.5 of Anderson (1971). ∎

4. Robustness in Mixed Regression and Autoregression

Now we consider the model

$$
(4.1) \qquad \boldsymbol{x}_t = \boldsymbol{B}\boldsymbol{x}_{t-1} + \boldsymbol{\Delta}\boldsymbol{z}_t + \boldsymbol{v}_t, \quad t = 1, 2, \ldots .
$$

This model is analogous to the regression model (2.1) with $\boldsymbol{z}_t$ replaced by $(\boldsymbol{x}'_{t-1}, \boldsymbol{z}'_t)'$. The least squares estimator of $(\boldsymbol{B}, \boldsymbol{\Delta})$ is

$$
(4.2) \qquad (\hat{\boldsymbol{B}}_n, \hat{\boldsymbol{\Delta}}_n) = \left(\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}'_{t-1}, \sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{z}'_t\right)\begin{bmatrix} \sum_{t=1}^{n}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1} & \sum_{t=1}^{n}\boldsymbol{x}_{t-1}\boldsymbol{z}'_t \\ \sum_{t=1}^{n}\boldsymbol{z}_t\boldsymbol{x}'_{t-1} & \sum_{t=1}^{n}\boldsymbol{z}_t\boldsymbol{z}'_t \end{bmatrix}^{-1}.
$$

and the estimator of $\boldsymbol{\Sigma}$ is

$$
(4.3) \qquad \hat{\boldsymbol{\Sigma}}_n = \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{x}_t - \hat{\boldsymbol{B}}_n\boldsymbol{x}_{t-1} - \hat{\boldsymbol{\Delta}}_n\boldsymbol{z}_t)(\boldsymbol{x}_t - \hat{\boldsymbol{B}}_n\boldsymbol{x}_{t-1} - \hat{\boldsymbol{\Delta}}_n\boldsymbol{z}_t)'.
$$

Theorem 7. Let $\mathcal{E}\boldsymbol{x}_0\boldsymbol{x}_0' = \boldsymbol{\Sigma}_0$; let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ be generated by (4.1), and let $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots$ be a sequence of random variables (possibly degenerate). Let $\{\mathcal{F}_t\}$ be a sequence of increasing σ fields such that $\boldsymbol{v}_t$ is $\mathcal{F}_t$-measurable and $\boldsymbol{z}_t$ is $\mathcal{F}_{t-1}$-measurable. Suppose the characteristic roots of $\boldsymbol{B}$ are less than 1 in absolute value, $\mathcal{E}(\boldsymbol{v}_t|\mathcal{F}_{t-1}) = \boldsymbol{0}$ a.s., $\mathcal{E}(\boldsymbol{v}_t\boldsymbol{v}_t'|\mathcal{F}_{t-1}) = \boldsymbol{\Sigma}_t$ a.s., and (2.5), (2.26), and (2.41) hold. Suppose

$$\frac{1}{n}\sum_{t=1}^{n-h} \boldsymbol{z}_{t+h}\boldsymbol{z}_t' \xrightarrow{\text{p}} \boldsymbol{M}_h = \boldsymbol{M}_{-h}', \quad h = 0, 1, 2, \ldots , \tag{4.4}$$

$$\frac{1}{n}\sum_{t=1}^{n-h} \boldsymbol{z}_{t+h}\boldsymbol{v}_t' \xrightarrow{\text{p}} \boldsymbol{0}, \quad h = 1, 2, \ldots . \tag{4.5}$$

Define

$$\boldsymbol{L} = \sum_{s=0}^{\infty} \boldsymbol{B}^s \boldsymbol{\Delta} \boldsymbol{M}_{-(s+1)}. \tag{4.6}$$

Then

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_{t-1}\boldsymbol{z}_t' \xrightarrow{\text{p}} \boldsymbol{L}. \tag{4.7}$$

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}' \xrightarrow{\text{p}} \boldsymbol{Q}, \tag{4.8}$$

where $\boldsymbol{Q}$ is the unique solution to

$$\boldsymbol{Q} - \boldsymbol{B}\boldsymbol{Q}\boldsymbol{B}' = \boldsymbol{\Sigma} + \boldsymbol{B}\boldsymbol{L}\boldsymbol{\Delta}' + \boldsymbol{\Delta}\boldsymbol{L}'\boldsymbol{B}' + \boldsymbol{\Delta}\boldsymbol{M}_0\boldsymbol{\Delta}'. \tag{4.9}$$

Furthermore, if (2.42) and (3.23) hold and

$$\frac{1}{n}\sum_{t=1}^{n} (\boldsymbol{\Sigma}_t \otimes \boldsymbol{v}_{t-1-s}\boldsymbol{z}_t') \xrightarrow{\text{p}} \boldsymbol{0}, \quad s = 1, 2, \ldots , \tag{4.10}$$

then

$$\sqrt{n}\,\text{vec}(\hat{\boldsymbol{B}}_n - \boldsymbol{B}, \hat{\boldsymbol{\Delta}}_n - \boldsymbol{\Delta}) \xrightarrow{\mathcal{L}} N\left[\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{pmatrix}, \begin{pmatrix}\boldsymbol{Q} & \boldsymbol{L}\\ \boldsymbol{L}' & \boldsymbol{M}_0\end{pmatrix}^{-1} \otimes \boldsymbol{\Sigma}\right]. \tag{4.11}$$

and (2.32) holds under the further assumption that the inverse matrix in (4.11) exists.

Proof. Because the roots of $\boldsymbol{B}$ are less than 1 in absolute value, the sum in (4.6) converges (by use of the Cauchy–Schwarz inequality). From (4.1) we obtain

$$\begin{aligned}
\boldsymbol{x}_{t-1} &= \sum_{s=0}^{t-2} \boldsymbol{B}^s \boldsymbol{v}_{t-1-s} + \boldsymbol{B}^{t-1}\boldsymbol{x}_0 + \sum_{s=0}^{t-2} \boldsymbol{B}^s \boldsymbol{\Delta z}_{t-1-s} \\
&= \sum_{s=0}^{k} \boldsymbol{B}^s \boldsymbol{v}_{t-1-s} + \sum_{s=k+1}^{\infty} \boldsymbol{B}^s \boldsymbol{v}_{t-1-s} + \boldsymbol{B}^{t-1}\boldsymbol{x}_0 \\
&\quad + \sum_{s=0}^{k} \boldsymbol{B}^s \boldsymbol{\Delta z}_{t-1-s} + \sum_{s=k+1}^{\infty} \boldsymbol{B}^s \boldsymbol{\Delta z}_{t-1-s}.
\end{aligned} \tag{4.12}$$

where $\boldsymbol{v}_0 = \boldsymbol{v}_{-1} = \cdots = \mathbf{0}$ and $\boldsymbol{z}_0 = \boldsymbol{z}_{-1} = \cdots = \mathbf{0}$. Then

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_{t-1}\boldsymbol{z}_t' &= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=0}^{k} \boldsymbol{B}^s(\boldsymbol{v}_{t-1-s} + \boldsymbol{\Delta z}_{t-1-s})\boldsymbol{z}_t' \\
&\quad + \frac{1}{n}\sum_{t=1}^{n}\left[\boldsymbol{B}^{t-1}\boldsymbol{x}_0\boldsymbol{z}_t' + \sum_{s=k+1}^{\infty} \boldsymbol{B}^s(\boldsymbol{v}_{t-1-s} + \boldsymbol{\Delta z}_{t-1-s})\boldsymbol{z}_t'\right].
\end{aligned} \tag{4.13}$$

We calculate by use of Lemma 7

$$\begin{aligned}
\left|\frac{1}{n}\sum_{t=1}^{n}\sum_{s=k+1}^{\infty} \boldsymbol{B}^s \boldsymbol{v}_{t-1-s}\boldsymbol{z}_t'\right| &\le \frac{1}{n}\sum_{t=1}^{n}\sum_{s=k+1}^{\infty} \lambda^s s^{p-1} q^{**}(\|\boldsymbol{v}_{t-1-s}\|^2 + \|\boldsymbol{z}_t\|^2) \\
&\le q^{**}\sum_{s=k+1}^{\infty} \lambda^s s^{p-1} \frac{1}{n}\sum_{t=1}^{n}(\|\boldsymbol{v}_t\|^2 + \|\boldsymbol{z}_t\|^2).
\end{aligned} \tag{4.14}$$

Since $\sum_{s=0}^{\infty} \lambda^s s^{p-1}$ converges and $\sum_{t=1}^{n} \|\boldsymbol{z}_t\|^2/n \xrightarrow{p} \operatorname{tr} \boldsymbol{M}_0$, we can choose k sufficiently large to make the right-hand side of (4.14) arbitrarily small with arbitrarily high probability. Similarly the other two terms in the second sum in (4.13) can be made small. Then

$$\frac{1}{n}\sum_{s=0}^{k} \boldsymbol{B}^s(\boldsymbol{v}_{t-1-s} + \boldsymbol{\Delta z}_{t-1-s})\boldsymbol{z}_t' \xrightarrow{p} \frac{1}{n}\sum_{s=0}^{k} \boldsymbol{B}^s \boldsymbol{\Delta M}_{-k}. \tag{4.15}$$

That leads to (4.7).

From (4.1) we have

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{v}_t\boldsymbol{v}_t' = \frac{1}{n}\sum_{t=1}^{n}\Big[\boldsymbol{x}_t\boldsymbol{x}_t' - \boldsymbol{B}\boldsymbol{x}_{t-1}\boldsymbol{x}_t' - \boldsymbol{\Delta z}_t\boldsymbol{x}_t' \tag{4.16}$$

$$
\begin{aligned}
&- \boldsymbol{x}_t\boldsymbol{x}'_{t-1}\boldsymbol{B}' + \boldsymbol{B}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\boldsymbol{B}' + \boldsymbol{\Delta}\boldsymbol{z}_t\boldsymbol{x}'_{t-1}\boldsymbol{B}' \\
&- \boldsymbol{x}_t\boldsymbol{z}'_t\boldsymbol{\Delta}' + \boldsymbol{B}\boldsymbol{x}_{t-1}\boldsymbol{z}'_t\boldsymbol{\Delta}' + \boldsymbol{\Delta}\boldsymbol{z}_t\boldsymbol{z}'_t\boldsymbol{\Delta}'\Big] \\
\xrightarrow{\text{P}}\ & \boldsymbol{\Sigma}.
\end{aligned}
$$

$$
(4.17) \qquad \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{v}_t\boldsymbol{x}'_{t-1} = \frac{1}{n}\sum_{t=1}^{n}\Big[\boldsymbol{x}_t\boldsymbol{x}'_{t-1} - \boldsymbol{B}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1} - \boldsymbol{\Delta}\boldsymbol{z}_t\boldsymbol{x}'_{t-1}\Big].
$$

$$
(4.18) \qquad \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{v}_t\boldsymbol{z}'_t = \frac{1}{n}\sum_{t=1}^{n}\Big[\boldsymbol{x}_t\boldsymbol{z}'_t - \boldsymbol{B}\boldsymbol{x}_{t-1}\boldsymbol{z}'_t - \boldsymbol{\Delta}\boldsymbol{z}_t\boldsymbol{z}'_t\Big] \xrightarrow{\text{P}} \boldsymbol{0}.
$$

If (4.17) $\xrightarrow{\text{P}} \boldsymbol{0}$, then from (4.16), (4.17), and (4.18) we obtain

$$
\begin{aligned}
(4.19) \qquad & \frac{1}{n}\sum_{t=1}^{n}\Big(\boldsymbol{x}_t\boldsymbol{x}'_t - \boldsymbol{B}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\boldsymbol{B}'\Big) \\
&= \frac{1}{n}\Big[\sum_{t=1}^{n}(\boldsymbol{x}_t\boldsymbol{x}'_t - \boldsymbol{B}\boldsymbol{x}_t\boldsymbol{x}'_t\boldsymbol{B}') + \boldsymbol{B}\boldsymbol{x}_n\boldsymbol{x}'_n\boldsymbol{B}' - \boldsymbol{B}\boldsymbol{x}_0\boldsymbol{x}'_0\boldsymbol{B}'\Big] \\
&\xrightarrow{\text{P}} \boldsymbol{\Sigma} + \boldsymbol{B}\boldsymbol{L}\boldsymbol{\Delta}' + \boldsymbol{\Delta}\boldsymbol{L}'\boldsymbol{B}' + \boldsymbol{\Delta}\boldsymbol{M}_0\boldsymbol{\Delta}'.
\end{aligned}
$$

If $(1/n)\boldsymbol{x}'_n\boldsymbol{x}_n \xrightarrow{\text{P}} 0$, then (4.8) follows from (4.19). Thus

$$
(4.20) \qquad \frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix}\boldsymbol{x}_{t-1}\\ \boldsymbol{z}_t\end{pmatrix}(\boldsymbol{x}'_{t-1}, \boldsymbol{z}'_t) \xrightarrow{\text{P}} \begin{pmatrix}\boldsymbol{Q} & \boldsymbol{L}\\ \boldsymbol{L}' & \boldsymbol{M}_0\end{pmatrix}.
$$

Now we consider

$$
\begin{aligned}
(4.21) \qquad & \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{\Sigma}_t \otimes \boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}) \\
&= \frac{1}{n}\sum_{t=1}^{n}\left[\boldsymbol{\Sigma}_t \otimes \sum_{r,s=0}^{\infty}\boldsymbol{B}^r(\boldsymbol{\Delta}, \boldsymbol{I})\begin{pmatrix}\boldsymbol{z}_{t-1-s}\\ \boldsymbol{v}_{t-1-s}\end{pmatrix}(\boldsymbol{z}'_{t-1-s}, \boldsymbol{v}'_{t-1-s})\begin{pmatrix}\boldsymbol{\Delta}'\\ \boldsymbol{I}\end{pmatrix}(\boldsymbol{B}')^s\right].
\end{aligned}
$$

If the sums in (4.21) on r, s run from $k+1$ to ∞, the trace converges to an arbitrarily small quantity by taking k sufficiently large. Then

$$
\begin{aligned}
(4.22) \qquad & \frac{1}{n}\sum_{t=1}^{n}\left[\boldsymbol{\Sigma}_t \otimes \sum_{r,s=0}^{k}\boldsymbol{B}^r(\boldsymbol{\Delta}, \boldsymbol{I})\begin{pmatrix}\boldsymbol{z}_{t-1-r}\\ \boldsymbol{v}_{t-1-s}\end{pmatrix}(\boldsymbol{z}'_{t-1-s}, \boldsymbol{v}'_{t-1-s})\begin{pmatrix}\boldsymbol{\Delta}'\\ \boldsymbol{I}\end{pmatrix}(\boldsymbol{B}')^s\right] \\
&\xrightarrow{\text{P}} \boldsymbol{\Sigma} \otimes \sum_{r,s=0}^{k}\boldsymbol{B}^r\big[\boldsymbol{\Delta}\boldsymbol{M}_{s-r}\boldsymbol{\Delta}' + \delta_{r,s}\boldsymbol{\Sigma}\big](\boldsymbol{B}')^s.
\end{aligned}
$$

Thus

$$(4.23)\qquad \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{\Sigma}_t \otimes \boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}) \xrightarrow{\text{p}} \boldsymbol{\Sigma} \otimes \left[\sum_{r,s=0}^{\infty} \boldsymbol{B}^r \boldsymbol{\Delta} \boldsymbol{M}_{s-r}\boldsymbol{\Delta}'(\boldsymbol{B}')^s + \sum_{s=0}^{\infty} \boldsymbol{B}^s \boldsymbol{\Sigma}(\boldsymbol{B}')^s\right]$$
$$= \boldsymbol{\Sigma} \otimes \boldsymbol{Q}.$$

By similar means we can complete the proof of

$$(4.24)\qquad \frac{1}{n}\sum_{t=1}^{n}\left[\boldsymbol{\Sigma}_t \otimes \begin{pmatrix}\boldsymbol{x}_{t-1}\\ \boldsymbol{z}_t\end{pmatrix}(\boldsymbol{x}'_{t-1}, \boldsymbol{z}'_t)\right] \longrightarrow \begin{pmatrix}\boldsymbol{Q} & \boldsymbol{L}\\ \boldsymbol{L}' & \boldsymbol{M}_0\end{pmatrix}.$$

Theorem 1 can then be applied with $\boldsymbol{z}_t$ in Theorem 1 replaced by $(\boldsymbol{x}'_{t-1}, \boldsymbol{z}'_t)'$ to obtain (4.11), and (2.33) follows.

To apply Theorem 1 we also need

$$(4.25)\qquad \frac{1}{n}\max_{t=1,\ldots,n}\|\boldsymbol{x}'_{t-1}\|^2 \xrightarrow{\text{p}} 0.$$

To prove this we need only consider

$$(4.26)\qquad \boldsymbol{x}^*_{t-1} = \sum_{s=0}^{t-2}\boldsymbol{B}^s(\boldsymbol{v}_{t-1-s} + \boldsymbol{\Delta}\boldsymbol{z}_{t-1-s}).$$

Then

$$(4.27)\qquad \boldsymbol{x}^{*\prime}_{t-1}\boldsymbol{x}^*_{t-1} = \left\|\sum_{s=0}^{t-2}\boldsymbol{B}^s(\boldsymbol{v}_{t-1-s} + \boldsymbol{\Delta}\boldsymbol{z}_{t-1-s}\right\|^2$$
$$\leq 2\left\|\sum_{s=2}^{t-2}\boldsymbol{B}^s\boldsymbol{v}_{t-1-s}\right\|^2 + 2\left\|\sum_{s=0}^{t-2}\boldsymbol{B}^s\boldsymbol{\Delta}\boldsymbol{z}_{t-1-s}\right\|^2.$$

By (3.4) the first term on the right-hand side of (4.27) is less than or equal to

$$(4.28)\qquad 4\sum_{r,s=0}^{t-2}\lambda^{r+s}qr^{p-1}s^{p-1}\|\boldsymbol{v}_{t-1-s}\|^2 \leq 4\sum_{r,s=0}^{t-2}\lambda^{r+s}qr^{p-1}s^{p-1}\max_{t=1,\ldots,n}\|\boldsymbol{v}_t\|^2.$$

Since $\|\boldsymbol{\Delta}\boldsymbol{z}_{t-1-s}\|^2 \leq \text{const}\,\|\boldsymbol{z}_{t-1-s}\|^2$, we obtain

$$(4.29)\qquad \|\boldsymbol{x}*_{t-1}\|^2 \leq 4\sum_{r,s=0}^{t-2}\lambda^{r+s}r^{p-1}s^{p-1}\left(q\max_{t=1,\ldots,n}\|\boldsymbol{v}_t\|^2 + q*\max_{t=1,\ldots,n}\|\boldsymbol{z}_t\|^2\right),$$

which implies (4.25) and $\|\boldsymbol{x}_n\|^2/n \xrightarrow{\text{p}} 0$.

Now we want to show that

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_{t-1}\boldsymbol{v}_t' \xrightarrow{\text{p}} \mathbf{0}. \tag{4.30}$$

From (4.12) we have

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_{t-1}\boldsymbol{v}_t' &= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=0}^{t-2} \boldsymbol{B}^s \boldsymbol{v}_{t-s-1}\boldsymbol{v}_t' \\ &\quad + \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{B}^{t-1}\boldsymbol{x}_0\boldsymbol{v}_t' + \frac{1}{n}\sum_{t=1}^{n}\sum_{s=0}^{t-2} \boldsymbol{B}^s \boldsymbol{\Delta z}_{t-s-1}\boldsymbol{v}_t'.\end{aligned} \tag{4.31}$$

It was shown in Section 3 that the first two terms on the right-hand side of (4.31) converge to $\mathbf{0}$ in probability as $n \to \infty$.

Define $\boldsymbol{v}_{nt}$ by (3.10) and $\boldsymbol{z}_{nt}$ by

$$\boldsymbol{z}_{nt} = \boldsymbol{z}_t I(\|\boldsymbol{z}_t\|^2 \le n). \tag{4.32}$$

Then

$$\Pr\left\{\frac{1}{n}\sum_{t=1}^{n}\sum_{s=0}^{t-2} \boldsymbol{B}^s \boldsymbol{\Delta z}_{t-s-1}\boldsymbol{v}_t' = \frac{1}{n}\sum_{t=1}^{n}\sum_{s=0}^{t-2} \boldsymbol{B}^s \boldsymbol{\Delta} z_{n,t-s-1}\boldsymbol{v}_{nt}\right\} \to 1. \tag{4.33}$$

Consider

$$\begin{aligned}&\mathcal{E}\,\mathrm{tr}\left(\frac{1}{n}\sum_{t=1}^{n}\sum_{s=0}^{t-2} \boldsymbol{B}^s \boldsymbol{\Delta z}_{n,t-s-1}\boldsymbol{v}_{nt}'\right)'\left(\frac{1}{n}\sum_{t=1}^{n}\sum_{s=0}^{t-s} \boldsymbol{B}^r \boldsymbol{\Delta z}_{n,t-r-1}\boldsymbol{v}_{nt}'\right) \\ &= \frac{1}{n^2}\mathcal{E}\left[\sum_{t=1}^{n}\left(\sum_{s=0}^{t-2} \boldsymbol{B}^s \boldsymbol{\Delta} z_{n,t-s-1}\right)'\left(\sum_{r=0}^{t-2} \boldsymbol{B}^r \boldsymbol{\Delta z}_{n,t-r-1}\right)\mathcal{E}(\boldsymbol{v}_{nt}'\boldsymbol{v}_{nt}|\mathcal{F}_{t-1})\right] \\ &= \frac{1}{n^2}\mathcal{E}\sum_{t=1}^{n}\|\sum_{s=0}^{t-2} \boldsymbol{B}^s \boldsymbol{\Delta z}_{n,t-s-1}\|^2\mathcal{E}(\boldsymbol{v}_{nt}'\boldsymbol{v}_{nt}|\mathcal{F}_{t-1}) \\ &\le \frac{1}{n}\mathcal{E}\max_{s=1,\dots,n}\|\boldsymbol{z}_{ns}\|^2\sum_{s=0}^{n-1}\mathrm{tr}\ \boldsymbol{\Delta}'(\boldsymbol{B}')^s\boldsymbol{B}^s\boldsymbol{\Delta}\mathcal{E}(\boldsymbol{v}_{nt}'\boldsymbol{v}_{nt}|\mathcal{F}_{t-1}) \\ &\to 0\end{aligned} \tag{4.34}$$

because $\|\boldsymbol{z}_{ns}\|^2/n \xrightarrow{\text{p}} 0$ and $\|\boldsymbol{z}_{ns}\|^2$ is bounded and $\boldsymbol{\Sigma}_t \xrightarrow{\text{p}} \boldsymbol{\Sigma}$ and $\|\boldsymbol{v}_{nt}\|^2$ is bounded. This proves (4.20) and the theorem. ∎

Lemma 6. If assumptions of Theorem 7 hold and if $\boldsymbol{\Sigma}$ and $\boldsymbol{M}_0$ are positive definite, then (4.24) is positive definite.

Proof.

$$(4.35) \qquad (\boldsymbol{c}', \boldsymbol{d}')\begin{pmatrix} \boldsymbol{Q} & \boldsymbol{L} \\ \boldsymbol{L}' & \boldsymbol{M}_0 \end{pmatrix}\begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{d} \end{pmatrix} = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{c}'\boldsymbol{x}_{t-1} + \boldsymbol{d}'\boldsymbol{z}_t)^2$$

$$= \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n}\Big[(\boldsymbol{c}'\boldsymbol{v}_{t-1})^2 + (\boldsymbol{c}'\boldsymbol{B}\boldsymbol{x}_{t-1} + \boldsymbol{c}'\boldsymbol{\Delta}\boldsymbol{z}_{t-1} + \boldsymbol{d}'\boldsymbol{z}_t)^2$$

$$+2\boldsymbol{c}'\boldsymbol{v}_{t-1}(\boldsymbol{x}'_{t-2}\boldsymbol{B}'\boldsymbol{c} + \boldsymbol{z}'_{t-1}\boldsymbol{\Delta}'\boldsymbol{c} + \boldsymbol{z}'_t\boldsymbol{d})\Big]$$

$$= \boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c} + \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{c}'\boldsymbol{B}\boldsymbol{x}_{t-1} + \boldsymbol{c}'\boldsymbol{\Delta}\boldsymbol{z}_{t-1} + \boldsymbol{d}'\boldsymbol{z}_t)^2$$

$$\geq \boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}$$

by (4.3) and (4.30). If the left-hand side of (4.35) is 0, then $\boldsymbol{c} = \boldsymbol{0}$ because $\boldsymbol{\Sigma}$ is positive definite. In that case the left-hand side of (4.35) is $\boldsymbol{d}'\boldsymbol{M}_0\boldsymbol{d} = \boldsymbol{0}$; since $\boldsymbol{M}_0$ is positive definite, $\boldsymbol{d} = \boldsymbol{0}$. ∎

A special case of the mixed model is $\boldsymbol{z}_t = 1$. Then (4.1) is

$$(4.36) \qquad \boldsymbol{x}_t = \boldsymbol{B}\boldsymbol{x}_{t-1} + \boldsymbol{\gamma} + \boldsymbol{v}_t,$$

where $\boldsymbol{\gamma} = \boldsymbol{\Delta}$ or

$$(4.37) \qquad \boldsymbol{x}_t - \boldsymbol{\mu} = \boldsymbol{B}(\boldsymbol{x}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{v}_t,$$

where $\boldsymbol{\gamma} = (\boldsymbol{I} - \boldsymbol{B})\boldsymbol{\mu}$. In this case (2.41), (4.4) and (4.5) are automatically satisfied, and condition (4.10) reduces to

$$(4.38) \qquad \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{\Sigma}_t \otimes \boldsymbol{v}_{t-1-s}) \xrightarrow{\mathrm{P}} \boldsymbol{0}, \quad s = 0, 1, \ldots .$$

The matrix $\boldsymbol{L}$ is

$$(4.39) \qquad \boldsymbol{L} = \sum_{s=0}^{\infty}\boldsymbol{B}^s\boldsymbol{\gamma} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\gamma},$$

and the matrix $\boldsymbol{Q}$ is

$$\boldsymbol{Q} = \boldsymbol{\Gamma} + (\boldsymbol{I} - \boldsymbol{B})^{-1}\gamma\gamma'(\boldsymbol{I} - \boldsymbol{B}')^{-1}. \tag{4.40}$$

In this case

$$\hat{\boldsymbol{B}}_n = \left(\sum_{t=1}^{n} \boldsymbol{x}_t\boldsymbol{x}'_{t-1} - \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t \sum_{t=1}^{n}\boldsymbol{x}'_{t-1}\right)\left(\sum_{t=1}^{n}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1} - \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_{t-1}\sum_{t=1}^{n}\boldsymbol{x}'_{t-1}\right)^{-1} \tag{4.41}$$

and $\hat{\boldsymbol{\mu}}_n = (\boldsymbol{I} - \hat{\boldsymbol{B}}_n)\hat{\gamma}_n$, which is approximately $(1/n)\sum_{t=1}^{n}\boldsymbol{x}_t$. The limiting covariance matrix of $\sqrt{n}\big[(1/n)\sum_{t=1}^{n}\boldsymbol{x}_t - \boldsymbol{\mu}\big]$ is

$$(\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Gamma} + \boldsymbol{\Gamma}(\boldsymbol{I} - \boldsymbol{B}')^{-1} - \boldsymbol{\Gamma}. \tag{4.42}$$

The condition (4.5) suggests a kind of lack of correlation between $\boldsymbol{z}_{t+h}$ and $\boldsymbol{v}_t$ which is plausible if $\{\boldsymbol{z}_t\}$ and $\{\boldsymbol{v}_t\}$ are independent; that is, if the $\boldsymbol{z}_t$'s are exogenous.

Acknowledgements. The authors are indebted to T. L. Lai for helpful suggestions. The research of the first author was supported by U.S. Army Research Office Contract DAAL03-89-K-0033; the research of the second author was supported by Grant-in-Aid 01301075 of the Ministry of Education at the Faculty of Economics, University of Tokyo, Japan.

Appendix

Lemma 7. Let the largest absolute value of the characteristic roots of $\boldsymbol{B}$ of order p be $\lambda < 1$. Then for any vectors $\boldsymbol{u}$ and $\boldsymbol{v}$

$$|\boldsymbol{u}'(\boldsymbol{B}')^r \boldsymbol{B}^s \boldsymbol{v}| \leq \lambda^{r+s} q r^{p-1} s^{p-1} (\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2) \tag{A.1}$$

for a suitable constant q.

Proof. There exists a matrix $\boldsymbol{P}$ such that $\boldsymbol{B} = \boldsymbol{P}^{-1}\boldsymbol{H}\boldsymbol{P}$, where

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{H}_K \end{bmatrix}. \tag{A.2}$$

the $p_k \times p_k$ matrix $\boldsymbol{H}_k = \lambda_k \boldsymbol{I} + \boldsymbol{L}_k$, λ_k is a characteristic root of $\boldsymbol{B}$, and

$$\boldsymbol{L}_k = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}. \tag{A.3}$$

Then

$$\boldsymbol{u}'(\boldsymbol{B}')^r \boldsymbol{B}^s \boldsymbol{v} = \boldsymbol{u}'\boldsymbol{P}'(\boldsymbol{H}')^r (\boldsymbol{P}\boldsymbol{P}')^{-1} \boldsymbol{H}^s \boldsymbol{P} \boldsymbol{v}. \tag{A.4}$$

Let

$$(\boldsymbol{P}\boldsymbol{P}')^{-1} = \boldsymbol{G} = \begin{bmatrix} \boldsymbol{G}_{11} & \boldsymbol{G}_{12} & \cdots & \boldsymbol{G}_{1K} \\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22} & \cdots & \boldsymbol{G}_{2K} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{G}_{K1} & \boldsymbol{G}_{K2} & \cdots & \boldsymbol{G}_{KK} \end{bmatrix}. \tag{A.5}$$

For $s \geq p_k - 1$ we have

$$\begin{aligned} \boldsymbol{H}_k^s &= \lambda_k^s \boldsymbol{I} + \lambda_k^{s-1}\binom{s}{1}\boldsymbol{L}_k + \cdots + \lambda_k^{s-(p_k-1)}\binom{s}{p_k-1}\boldsymbol{L}_k^{p_k-1} \\ &= \lambda_k^s\left[\boldsymbol{I} + \lambda_k^{-1}\binom{s}{1}\boldsymbol{L}_k + \cdots + \lambda_k^{-(p_k-1)}\binom{s}{p_k-1}\boldsymbol{L}_k^{p_k-1}\right], \end{aligned} \tag{A.6}$$

$$(A.7) \qquad (\boldsymbol{H}_k')^r \boldsymbol{G}_{k\ell} \boldsymbol{H}_\ell^s = \lambda_k^r \lambda_\ell^s \Big[\boldsymbol{G}_{k\ell} + \lambda_k^{-1} \binom{r}{1} \boldsymbol{L}_k' \boldsymbol{G}_{k\ell} + \lambda_\ell^{-1} \binom{s}{1} \boldsymbol{G}_{k\ell} \boldsymbol{L}_\ell + \cdots$$

$$+ \lambda_k^{-r} \lambda_\ell^{-s} \binom{r}{p_{k-1}} \binom{s}{p_{\ell-1}} (\boldsymbol{L}_k')^{p_k - 1} \boldsymbol{G}_{k\ell} \boldsymbol{L}_\ell^{p_\ell - 1} \Big]$$

$$= \lambda_k^r \lambda_\ell^s \boldsymbol{Q}_{k\ell}(r,s).$$

Let $\boldsymbol{Pu} = \boldsymbol{x}$, $\boldsymbol{Pv} = \boldsymbol{y}$ and

$$(A.8) \qquad \boldsymbol{Q}(r,s) = \begin{bmatrix} \boldsymbol{Q}_{11}(r,s) & \boldsymbol{Q}_{12}(r,s) & \cdots & \boldsymbol{Q}_{1K}(r,s) \\ \boldsymbol{Q}_{21}(r,s) & \boldsymbol{Q}_{22}(r,s) & \cdots & \boldsymbol{Q}_{2K}(r,s) \\ \vdots & \vdots & & \vdots \\ \boldsymbol{Q}_{K1}(r,s) & \boldsymbol{Q}_{K2}(r,s) & \cdots & \boldsymbol{Q}_{KK}(r,s) \end{bmatrix}$$

$$= (q_{ij}(r,s)).$$

The element $q_{ij}(r,s)$ is a polynomial in r and s of degree at most $p-1$ with fixed coefficients. Then

$$(A.9) \qquad |\boldsymbol{x}' \lambda^{r+s} \boldsymbol{Q}(r,s) \boldsymbol{y}| \leq \lambda^{r+s} \sum_{i,j=1}^{p} |q_{ij}(r,s)| |x_i| |y_j|$$

$$\leq \lambda^{r+s} \sum_{i,j=1}^{p} \frac{|q_{ij}(r,s)|}{2} (x_i^2 + y_i^2)$$

$$\leq p \lambda^{r+s} \max_{i,j=1,\ldots,p} \frac{|q_{ij}(r,s)|}{2} (\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2).$$

Let

$$(A.10) \qquad q_{ij}(r,s) = \sum_{g,h=0}^{p-1} q_{ij}^{gh} r^g s^h.$$

Then

$$(A.11) \qquad \max_{i,j=1,\ldots,p} |q_{ij}(r,s)| \leq \max_{i,j=1,\ldots,p} \sum_{g,h=0}^{p-1} |q_{ij}^{gh}| r^{p-1} s^{p-1}$$

and $\|\boldsymbol{x}\|^2 \leq \|\boldsymbol{u}\|^2$ times the maximum characteristic root of $\boldsymbol{PP}'$ and similarly for $\|\boldsymbol{y}\|^2$. The lemma follows. ∎

Lemma 8. (3.28).

Proof. The left–hand side of (3.28) is positive semidefinite. Its trace is

$$\frac{1}{n}\sum_{t=1}^{n} \operatorname{tr} \boldsymbol{\Sigma}_t \operatorname{tr} \boldsymbol{x}_0'(\boldsymbol{B}')^{t-1}\boldsymbol{B}^{t-1}\boldsymbol{x}_0 \le \frac{1}{n}\sum_{t=1}^{n} \operatorname{tr} \boldsymbol{\Sigma}_t \lambda^{2t-2}t^{2p-2}q^*\|\boldsymbol{x}_0\|^2. \tag{A.12}$$

We can take t_0 large enough so that for $t > t_0$ and arbitrary $\varepsilon > 0, \delta > 0$

$$\Pr\{\lambda^{2t-2}t^{2p-2}q^*\|\boldsymbol{x}_0\|^2 < \varepsilon\} > 1 - \delta. \tag{A.13}$$

Then the right–hand side of (A.12) is with probability greater than $1 - \delta$ not greater than

$$\frac{1}{n}\sum_{t=1}^{n_0} \operatorname{tr} \boldsymbol{\Sigma}_t \lambda^{2t-2}t^{2p-2}q^*\|\boldsymbol{x}_0\|^2 + \varepsilon\frac{1}{n}\sum_{t=n_0}^{n} \operatorname{tr} \boldsymbol{\Sigma}_t \xrightarrow{\mathrm{P}} \varepsilon \operatorname{tr} \boldsymbol{\Sigma} \tag{A.14}$$

as $n \to \infty$. ∎

Comments on Condition (2.5)

A key assumption is

$$\sup_{t=1,2,\ldots} \mathcal{E}\left[\boldsymbol{v}_t'\boldsymbol{v}_t I(\boldsymbol{v}_t'\boldsymbol{v}_t > a)|\mathcal{F}_{t-1}\right] \xrightarrow{\mathrm{P}} 0 \tag{A.15}$$

as $a \to \infty$; that is, given $\varepsilon > 0$, $\delta > 0$ there exists a_0 such that for $a > a_0$

$$\Pr\left\{\sup_{t=1,2,\ldots} \mathcal{E}\left[\boldsymbol{v}_t'\boldsymbol{v}_t I(\boldsymbol{v}_t'\boldsymbol{v}_t > a)|\mathcal{F}_{t-1}\right] \le \varepsilon\right\} \ge 1 - \delta. \tag{A.16}$$

Let $W_t(a) = \mathcal{E}[\boldsymbol{v}_t'\boldsymbol{v}_t I(\boldsymbol{v}_t'\boldsymbol{v}_t > a)|\mathcal{F}_{t-1}]$. The above event for fixed a is

$$\bigcap_{t=1}^{\infty}\{W_t(a) \le \varepsilon\}, \tag{A.17}$$

which is measurable. The random variable

$$X_n(a) = \max_{t=1,\ldots,n} W_t(a) \tag{A.18}$$

has the property

$$X_{n+1}(a) = \max\left[X_n(a), W_{n+1}(a)\right]. \tag{A.19}$$

Note that for given a $X_n(a)$ is nondecreasing in n. The event (A.17) is

$$\left\{\lim_{n\to\infty} X_n(a) \le \varepsilon\right\} = \bigcap_{n=1}^{\infty}\left\{X_n(a) \le \varepsilon\right\}. \tag{A.20}$$

Note that since $X_n(a)$ can be defined by (A.19), it is a one–dimensional variable; that is, the condition is a weak condition. not a strong condition. It is a condition on the cdf's of $X_n(a)$.

References

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, John Wiley and Sons, Inc., New York.

Anderson, T. W. (1959). On the asymptotic distribution of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics*, **30**, 676-687.

Chan, N. H., and C. Z. Wei (1987). Asymptotic inference in nearly non-stationary AR(1) processes. *Annals of Statistics*, **15**, 1050-1063.

Chow, Y. W., and Henry Teicher (1988). *Probability Theory: Independence, Interchangeability, Martingales* (Second Edition). Springer-Verlag, New York.

Dvoretzky, Aryeh (1972). Asymptotic normality for sums of dependent random variables. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 2, University of California Press, Berkeley and Los Angeles, 513-535.

Hall, P., and C. C. Heyde (1980). *Martingale Limit Theory and Its Applications*. Academic Press, New York.

Lai, Tse-Leung, and Herbert Robbins (1981). Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes. *Zeitschrift für Wahrscheinlichkeitstheorie*, **56**, 329-360.

Lai, Tse-Leung, and David Siegmund (1983). Fixed accuracy estimation of an autoregressive parameter. *Annals of Statistics*, **11**, 478-485.

Lai, Tse-Leung, and C. Z. Wei (1983). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics*, **10**, 154-166.

Lindeberg, J. W. (1922). Eine neue Herleitung des exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, **15**, 211-225.

Mann, H. B., and A. Wald (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica*, **11**, 173-220.

Technical Reports
U.S. Army Research Office
Contracts DAAG29-82-K-0156, DAAG29-85-K-0239, and DAAL03-89-K-0033

1. "Maximum Likelihood Estimators and Likelihood Ratio Criteria for Multivariate Elliptically Contoured Distributions," T. W. Anderson and Kai-Tai Fang, September 1982.
2. "A Review and Some Extensions of Takemura's Generalizations of Cochran's Theorem," George P.H. Styan, September 1982.
3. "Some Further Applications of Finite Difference Operators," Kai-Tai Fang, September 1982.
4. "Rank Additivity and Matrix Polynomials," George P.H. Styan and Akimichi Takemura, September 1982.
5. "The Problem of Selecting a Given Number of Representative Points in a Normal Population and a Generalized Mills' Ratio," Kai-Tai Fang and Shu-Dong He, October 1982.
6. "Tensor Analysis of ANOVA Decomposition," Akimichi Takemura, November 1982.
7. "A Statistical Approach to Zonal Polynomials," Akimichi Takemura, January 1983.
8. "Orthogonal Expansion of Quantile Function and Components of the Shapiro-Francia Statistic," Akimichi Takemura, January 1983.
9. "An Orthogonally Invariant Minimax Estimator of the Covariance Matrix of a Multivariate Normal Population," Akimichi Takemura, April 1983.
10. "Relationships Among Classes of Spherical Matrix Distributions," Kai-Tai Fang and Han-Feng Chen, April 1984.
11. "A Generalization of Autocorrelation and Partial Autocorrelation Functions Useful for Identification of ARMA(p,q) Processes," Akimichi Takemura, May 1984.
12. "Methods and Applications of Time Series Analysis Part II: Linear Stochastic Models," T. W. Anderson and N. D. Singpurwalla, October 1984.
13. "Why Do Noninvertible Estimated Moving Averages Occur?" T. W. Anderson and Akimichi Takemura, November 1984.
14. "Invariant Tests and Likelihood Ratio Tests for Multivariate Elliptically Contoured Distributions," Huang Hsu, May 1985.
15. "Statistical Inferences in Cross-lagged Panel Studies," Lawrence S. Mayer, November 1985.
16. "Notes on the Extended Class of Stein Estimators," Suk-ki Hahn, July 1986.
17. "The Stationary Autoregressive Model," T. W. Anderson, July 1986.
18. "Bayesian Analyses of Nonhomogeneous Autoregressive Processes," T. W. Anderson, Nozer D. Singpurwalla, and Refik Soyer, September 1986.
19. "Estimation of a Multivariate Continuous Variable Panel Model," Lawrence S. Mayer and Kun Mao Chen, June 1987.
20. "Consistency of Invariant Tests for the Multivariate Analysis of Variance," T. W. Anderson and Michael D. Perlman, October 1987.
21. "Likelihood Inference for Linear Regression Models," T. J. DiCiccio, November 1987.
22. "Second-order Moments of a Stationary Markov Chain and Some Applications," T. W. Anderson, February 1989.
23. "Asymptotic Robustness in Regression and Autoregression Based on Lindeberg Conditions," T. W. Anderson and Naoto Kunitomo, June 1989.